

Phylogenetic Correlation Between Cellular Metabolism and Oncogenesis

By

Shayan Moghimyfiroozabad

Supervisor: Dr. Kambiz Nassirpour Alavian

This thesis is submitted to the Department of Biology as a partial fulfillment of the MSc
program in Neuroscience.

December 2016

Department of Biology

Faculty of Science

Baha'i Institute for Higher Education (BIHE)

© Shayan Moghimyfiroozabad, Tehran, Iran, 2016

Summary

One of the hallmarks of cancer cells is the reprogramming of metabolism, particularly glucose metabolism including glycolysis, TCA cycle, and oxidative phosphorylation (OXPHOS) in order to provide metabolites and energy for their growth and proliferation. Some of the most common oncogenes and tumor suppressors, including TP53, PTEN, PIK3CA, EGFR, and BRAF, are involved in regulation of cancer metabolism. To determine the oncogenic correlates of cancer metabolism, we performed a comprehensive analysis of the phylogenetic profiles of these oncogenes (here called glycolysis-, TCA cycle- and OXPHOS-related or GTO-Oncogenes), other oncogenes not implicated in metabolic reprogramming of cancer cells, and genes involved in glucose metabolism. Phylogenetic profiling of GTO-Oncogenes showed that they could be categorized into 3 subgroups according to their occurrences in different kingdoms and phyla. Subgroup 1, including TP53, was found predominantly in animals. Subgroup 2, including PIK3CA and PTEN, was found in eukaryotes, and Subgroup 3, including EGFR and BRAF, were found in both prokaryotes and eukaryotes. A comparison of the phylogenetic profiling comparison of these oncogenes with those of the glucose metabolism genes demonstrated that TP53 was highly correlated with the subunits of the complexes I, IV, and V of mitochondrial electron transport chain. The phylogenetic profile of PIK3CA displayed high correlation with some components of complexes I, III and IV of the electron transport chain as well as the hexokinase family involved in glycolysis. No significant correlation was observed between phylogenetic profiles of PTEN, EGFR, and BRAF with those of the glucose metabolic genes. In this study, we also predicted the function of other oncogenes in reprogramming of cancer cell metabolism by comparing their phylogenetic profiles with the GTO-Oncogenes profiles. Overall, our results suggest specific evolutionary connections and functional interactions between the process of oncogenic transformation of tumor cells and cellular metabolism.

Table of Contents

Summary	ii
Table of Contents	iii
List of Figures	v
List of Tables	vi
List of Abbreviations	vii
Acknowledgements	ix
Introduction	1
Hallmarks of Cancer	2
Apoptosis and Cancer Development	3
Cancer Development and Evolution	7
Non-Malignant Cells Expressing Warburg Effect	10
The importance of Oncogenes in Glucose Metabolism	11
<i>PI3K/Akt/mTOR</i>	11
<i>c-Myc</i>	13
<i>HIF1</i>	13
<i>Ras</i>	16
<i>BRAF</i>	17
<i>TP53</i>	17
Phylogenetic Profiling	20
Overall Hypothesis and Objectives	25

Methods and Materials	25
<i>Selection of Oncogenes</i>	<i>25</i>
<i>Selection of GTO-Oncogenes</i>	<i>26</i>
<i>Selection of Genes of Glucose Metabolism</i>	<i>26</i>
<i>Construction of Phylogenetic Profiles</i>	<i>27</i>
 Results	 27
<i>The Phylogenetic Profile of TP53 is Highly Correlated with Mitochondrial Enzymes</i>	<i>31</i>
<i>The Phylogenetic Profiles of PIK3CA is Correlated with Complexes 1, 3, and 4 of Mitochondrial ETC, as well as Hexokinases of Glycolysis Pathway.....</i>	<i>32</i>
<i>There's no Significant Correlation between the Phylogenetic Profiles of BRAF and EGFR with Components of Glycolysis/Gluconeogenesis, TCA, and OXPHOS</i>	<i>33</i>
<i>Phylogenetic Profiles of Random Genes found in Animals and Eukaryotes Have a Good Chance to Share Their Highly Correlated Genes of Glucose Metabolism with TP53 and PIK3CA</i>	<i>34</i>
<i>Novel connections between oncogenes and glucose metabolism.....</i>	<i>39</i>
 Discussion and Conclusion	 43
 References	 51

List of Figures

Figure 1. Differences of Glucose Metabolism in Normal versus Malignant Cells	5
Figure 2. The Characteristics of Cancer	6
Figure 3. Cytoplasmic Hybrids (Cybrids) Studies	8
Figure 4. Cancer Originates from Mitochondrial Impairment	10
Figure 5. A Signaling Network to Mediate Metabolic Shift in Cancer Cells Leading to Warburg Effect	15
Figure 6. The AMPK-TP53 Pathway	19
Figure 7. A Schematic Illustration of Phylogenetic Profiling Method of Binary Vectors	24
Figure 8. Categorization of the Oncogenes Involved in Metabolic Shift of Glycolysis, TCA Cycle and OXPHOS in Cancer Cells (GTO-Oncogenes)	29
Figure 9. Formation of Clusters in Glucose Metabolic Genes found in Prokaryotes and Eukaryotes	34
Figure 10. Correlation Matrix Plot of 87 Selected Oncogenes and Representatives of Subgroups 1, 2 & 3 of GTO-Oncogenes	42

List of Tables

Table 1. Classification of GTO-Oncogenes based on Their Mutation Rate 31

Table 2. Classification of 45 Candidate Oncogenes Based on Their Mutation Rate 45

List of Abbreviations

A

Ac-CoA: Acetyl-Coenzyme A
ACL: ATP Citrate Lyase
AMP: Adenosine Monophosphate
AMPK: AMP-activated Protein Kinase
APC: Adenomatous Polyposis Coli
ATP: Adenosine Triphosphate

B

BLAST: Basic Local Alignment Search Tool

C

Cit: Citrate
Cybrids: Cytoplasmic Hybrids

E

EGFR: Epidermal Growth Factor Receptor
ERK: Extracellular signal-Regulated Kinase
ETC: Electron Transport Chain

F

FAS: Fatty Acid Synthase

G

Glu: Glucose
GLUT1: Glucose Transporter 1
GRP78: Glucose Regulated Protein 78
GSH: Glutathione
GSK: Glycogen Synthase Kinase
GTO-Oncogenes: Glycolytic-TCA-OXPHOS-Correlated Oncogenes

H

HIF1: Hypoxia-inducible Factor 1
HMGCL: 3-Hydroxy-3-Methylglutaryl-CoA Lyase

I

IGF-1: Insulin-like Growth Factor 1

K

KEGG: Kyoto Encyclopedia of Genes and Genomes database

L

Lac: Lactate
LDH-A: Lactate Dehydrogenase A
LKB1: Liver serine/threonine Kinase B1

M

Mal: Malate
MAPK: Mitogen-Activated Protein Kinase
MDH: Malate Dehydrogenase
ME: Malic Enzyme
mTOR: mammalian (mechanistic) Target of Rapamycin

N

NADPH: Nicotinamide Adenine Dinucleotide Phosphate

O

OAA: Oxaloacetate
OXPHOS: Oxidative Phosphorylation

P

PDH: Pyruvate Dehydrogenase
PDK2: Pyruvate Dehydrogenase Kinase 2
PFK: Phosphofructokinase
PI3K: Phosphoinositide 3-Kinase
PIP3: Phosphatidylinositol (3,4,5) Triphosphate
PKA: cAMP-dependent Protein Kinase A
PTEN: Phosphatase and TENsin homolog
Pyr: Pyruvate

R

RHT: Restricted Hypothesis Testing

S

SCO2: Cytochrome C Oxidase subunit 2
SLP: Substrate Level Phosphorylation
SNP: Sodium Nitroprusside
SOD2: Superoxide Dismutase 2
Strap: Stress-responsive Activator of p300

STAT3: Signal Transducer and Activator of Transcription 3

T

TCA: Tricarboxylic Acid

TIGAR: TP53-Induced Glycolysis and Apoptosis Regulator

TP53: Tumor Suppressor Protein p53

V

VEGF: Vascular Endothelial Growth Factor

VHL: von Hippel-Lindau

Acknowledgements

I should thank my dearest family, my beloved supervisor, Dr. Kambiz Nassirpour Alavian who suggested the topic and hypothesis, and supervised the thesis with great attention and all my teachers and instructors who supported me emotionally and spiritually.

Introduction

Rapid and uncontrolled cell division and growth are the features of cancer cells. To accommodate the growth, cancer cells need to reprogram their metabolism according to their energy and material's requirements (Bost et al., 2016). Because of this feature of cancer cells, cancer has been categorized as a metabolic disease (Seyfried et al., 2014). In the presence of oxygen, normal cells obtain energy in the form of Adenosine Triphosphate (ATP) through reactions of glycolysis and electron transport chain (ETC). These metabolic reactions occur in cytoplasm and mitochondria, respectively, with higher efficiency for ATP production (~18x) from the ETC (Figure 1) (Bost et al., 2016). About 90 years ago, Otto Warburg discovered that the oxygen level in tumor cells has little effect on their respiration. He found both glycolysis (termed "fermentation" by Warburg) and respiration should be blocked in order to kill the cancer cells, since both of these energy-producing processes could fuel the cancer cells equally. These observations were in contrast to the normal cell, in which respiration is the main ATP generating process (Warburg et al., 1927, Warburg, 1956). The effect ascribed to the difference between the metabolism of normal and cancer cells is called "aerobic glycolysis" or "Warburg Effect". According to the "Warburg Effect", the glucose uptake and consumption of cancer cells are elevated to produce more ATP through glycolysis, while their reliance on energy obtained from mitochondrial oxidative phosphorylation (OXPHOS) is decreased even in the presence of oxygen (Figure 1) (Bost et al., 2016). Warburg considered this effect as a metabolic feature of all proliferating tumor cells. He thought that this effect occurs because cancer cells lose their functional mitochondria, resulting in an irreversible impairment of respiration (Warburg, 1956). In contrast to his original hypothesis, it has been shown that cancer cells require functional mitochondria for purposes other than ATP production, including intermediate metabolites. These metabolites, which are mostly obtained from tricarboxylic acid cycle (TCA), enable cancer cells to proliferate rapidly (Bost et al., 2016). For example, α -ketoglutarate and oxaloacetate are used for biosynthesis of nucleic acids, proteins, and lipids. Glutamine, the most plentiful amino acids in mammals, is another fuel for synthesis of TCA intermediates. Proliferating cells oxidize glutamine via TCA cycle to generate lactate and Nicotinamide Adenine Dinucleotide Phosphate (NADPH). This analogous process to aerobic glycolysis is called glutaminolysis (DeBerardinis et al., 2008). In

addition to the TCA cycle, it is suggested that intermediate metabolites of glycolysis are essential for biosynthesis of lipids and nucleotides through the pentose phosphate pathway (Weinberg et al., 2010). The pentose phosphate pathway has two arms: oxidative and non-oxidative. They generate ribose-5-phosphate and NADPH required for synthesis of nucleotide and fatty acids, respectively (Jones and Thompson, 2009). It seems that in tumor cells, mitochondria could adapt to new environmental conditions and availability of materials. Thus, instead of losing their classical roles in bioenergetics and becoming dysfunctional, they can gain novel functions by utilizing unusual strategies (e.g. reversing TCA). Subsequently, novel theories are emerging concluding that the cancer-adapted mitochondria have growth advantages for malignant cells (Hall et al., 2013). Mitochondria have, therefore, been suggested as therapeutic targets for cancer.

Hallmarks of Cancer:

Cancer is a disease resulting from genomic instabilities. These abnormalities generate tumor suppressor genes with recessive loss of function and oncogenes with dominant gain of function traits. Most and perhaps all kinds of human cancers are characterized by the following acquired molecular, cellular, and biochemical features: self-acquisition of growth, endless proliferation, immortalization and resistance to programmed cell death (apoptosis), lack of sensitivity to antiproliferative/growth-inhibitor signals, induction and maintenance of new blood vessels generation (angiogenesis), and metastasis and tissue invasion (Figure 2A) (Hanahan and Weinberg, 2000). These six hallmarks of cancer are acquired during various stages of tumor development. Tumor-promoting inflammation, genomic instability, and mutability are the fundamental features of cancer cells required for acquisition of these six hallmarks. Genomic instability results in scarce genetic alterations and mutations enabling the acquisition of cancer hallmarks and promoting tumorigenesis. The second enabling feature, inflammatory responses of immune system, could prompt proliferative or survival factors and other bioactive means for cancer cells to support them in acquisition of cancer characteristics (Hanahan and Weinberg, 2011).

Recently, two other hallmarks of cancer have been suggested: Immuno-evasion, i.e. avoiding destructive responses of immune system, and reprogramming of energy metabolism. These are two emerging features added to the aforementioned acquired capabilities of cancers. The first phenomenon results in the evasion of malignant cells from destructive responses of immune cells, especially macrophages, natural killer cells, and lymphocytes. This emerging hallmark, as well as the enabling feature of tumor-promoting inflammation, reveal the two opposing effects of the immune system in prevention or promotion of tumorigenesis. The other emerging feature, deregulation of energy metabolism, results in modifications of cellular energetics in cancer cells. Due to these metabolic alterations, tumor cells can provide sufficient bioactive molecules as well as energy for their cell proliferation and growth. These modifications result in increasing glucose uptake and glycolysis in cancer cells, even in the presence of oxygen, leading to synthesis of amino acids, nucleotides, and subsequently organelles. This metabolic deregulation is called Warburg effect, and it could be observed in other rapidly proliferating (non-cancerous) cells such as embryonic stem cells. As some of the oncogenes, including Myc or Ras, responsible for regulation of many of the acquired hallmarks of cancer, are activated during the reprogramming of energy metabolism, it is thought that this emerging characteristic could be an observable sign of the interaction and harmony of other cancer hallmarks (Hanahan and Weinberg, 2011).

Apoptosis and Cancer Development:

One of the important causes of cancer development is impaired apoptosis. The regulation of apoptosis is critically mediated by members of the Bcl-2 protein family. They contribute in immunity, homeostasis, development and tumor suppression. Various cellular stresses could activate the BH3-only members of the Bcl-2 family to trigger apoptosis. Upon activation, these proteins, such as Bim or Puma, bind and suppress their pro-survival (anti-apoptotic) relatives, including Bcl-2, Bcl-w, Bcl-x(L), A1, and Mcl-1, resulting in disinhibition and activation of pro-apoptotic members, Bak and Bax. Subsequently, this process leads to a succession of pro-apoptotic processes including disruption and permeabilization of the mitochondrial outer

membrane, release of pro-apoptotic proteins such as cytochrome c, and activation of caspases to promote apoptosis. Thus, imbalance in the expression and mitochondrial function of Bcl-2 proteins (e.g. decrease of a pro-apoptotic member or increase of an anti-apoptotic protein) could lead to cancer development (Willis and Adams, 2005, Adams and Cory, 2007, Hanahan and Weinberg, 2011). There are also other pathways containing oncogenes/tumor suppressors that contribute to apoptosis, such as Ras, RAF, Myc, and TP53. During cancer development, malignant cells may face stressful conditions that may cause programmed cell death. The signaling triggered by some oncogenic factors help these cells confront this natural barrier of tumorigenesis (Hanahan and Weinberg, 2011).

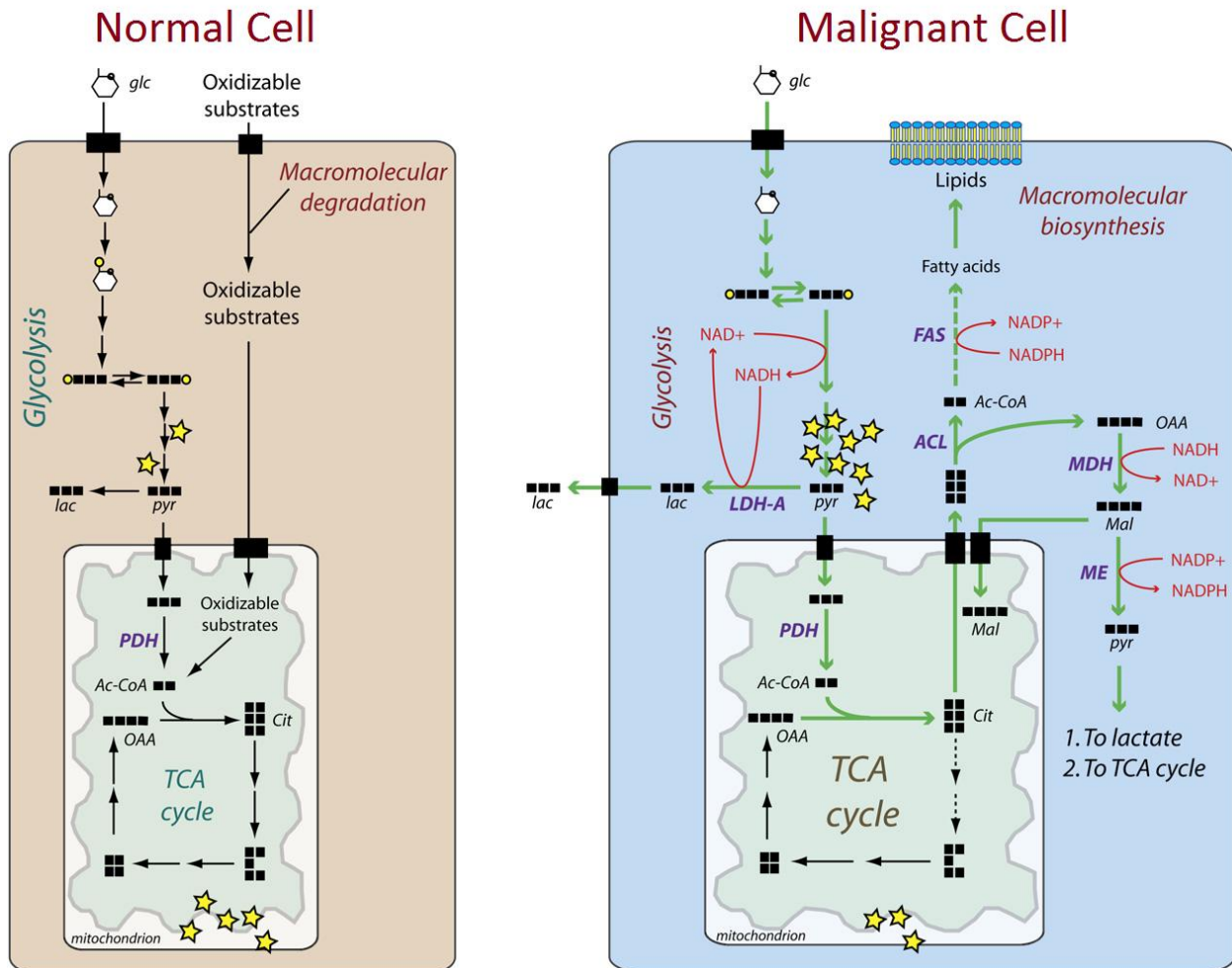


Figure 1. Differences of Glucose Metabolism in Normal versus Malignant Cells.

In normal cells, more ATP is obtained from oxidative phosphorylation of glucose (glu) or other oxidizable substrates such as fatty acids or amino acids. In malignant cells, glycolysis generates more energy in comparison with mitochondrial reactions. Instead, mitochondria function to provide prepare metabolic intermediates for biosynthesis of macromolecules.

glu: Glucose, pyr: Pyruvate, lac: Lactate, Cit: Citrate, Ac-CoA: Acetyl-CoA, OAA: Oxaloacetate, Mal: Malate, PDH: pyruvate dehydrogenase, LDH-A: lactate dehydrogenase A, ACL: ATP citrate lyase, FAS: fatty acid synthase, MDH: malate dehydrogenase, ME: malic enzyme.

This figure is taken from (DeBerardinis et al., 2008).

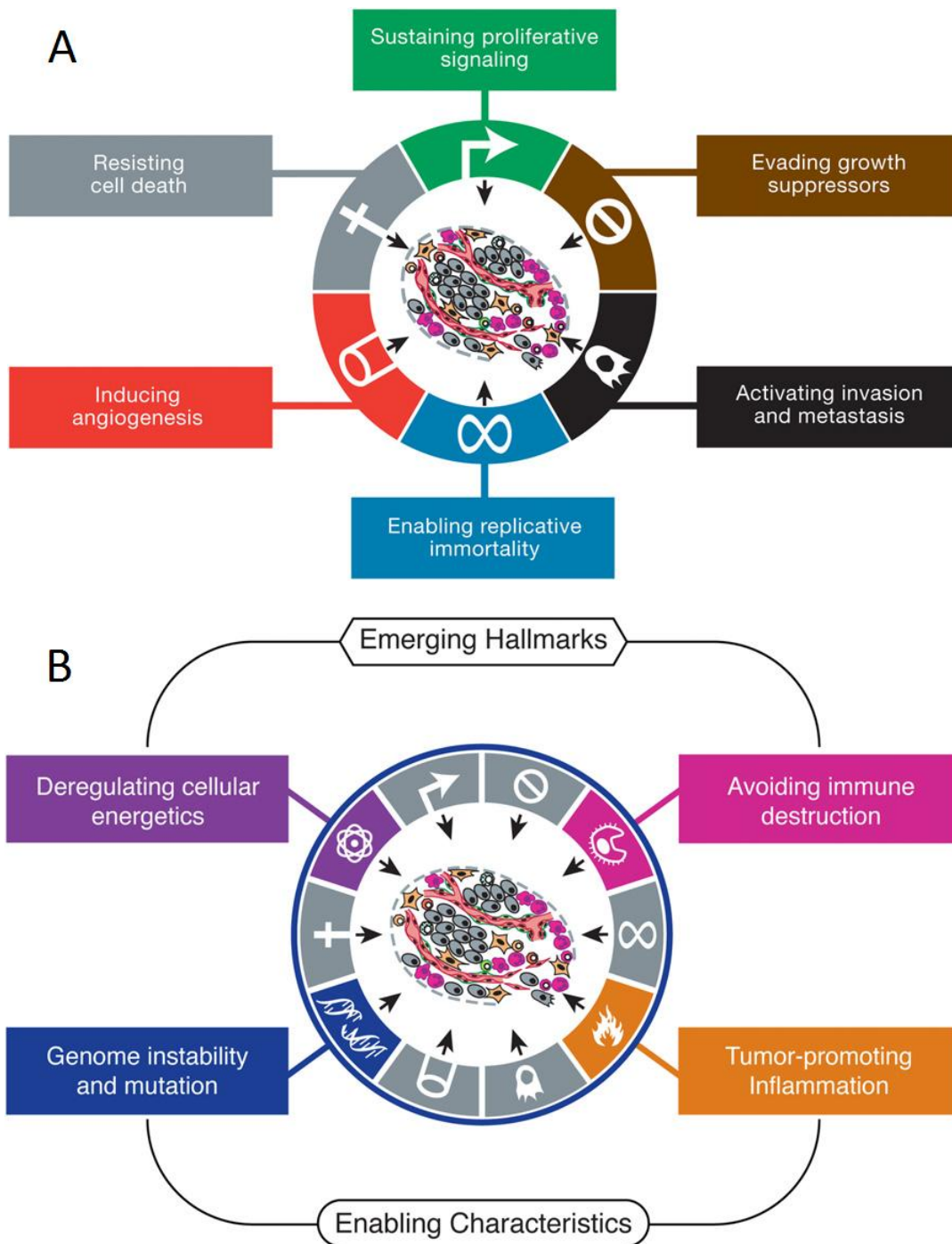


Figure 2. The Characteristics of Cancer.

A. The Six Acquired Hallmarks and **B.** The Enabling Characteristics & Emerging Hallmarks of Cancer.

This figure is taken from (Hanahan and Weinberg, 2011).

Cancer Development and Evolution:

Transition of a normal cell to a cancerous cell occurs in a multi-step neoplastic transformation process. During transformation, the cell obtains the aforementioned hallmarks of cancer preparing it to become malignant (Hanahan and Weinberg, 2011). This process is analogous to Darwinian evolution, in which alteration of normal cells to tumor cells is resulted from a sequence of genomic mutations. Each genomic alteration gives rise to one of the molecular, cellular and biochemical hallmarks of cancers representing as growth advantages (Hanahan and Weinberg, 2000), and driving forces for cancer progression (Seyfried et al., 2014).

There are some opposing views to the view of the genetic origin of cancer. They state that cancer has cytoplasmic causes. For example, Darlington suggested that there are particles in cytoplasm whose defects, but not nuclear genome's, cause tumorigenesis. These self-replicating particles which are susceptible to mutation are called "Plasmagenes" by Darlington. According to Plasmagene hypothesis, the transmission of these cytoplasmic particles does not obey the Mendelian rule, and does not show any recessive or dominant manner in their inheritance (Darlington, 1948). Although the nature of the plasmagene was not determined by Darlington, Seyfried *et al.* suggested that they could be mitochondria (Seyfried et al., 2014). A series of these observations were obtained from nuclear/cytoplasmic transfer studies between non-malignant and malignant cells to prepare cytoplasmic hybrids (cybrids). These studies show that when the cytoplasm of a non-malignant cell with normal mitochondria, was joined the nucleus of a malignant cell, the cancer development was suppressed or decreased significantly, and the combined cell transformed into or generated normal cells in spite of having genetic mutations (Israel and Schaeffer, 1987, Shay and Werbin, 1988, Li et al., 2003, Seyfried et al., 2014). On the other side, it was observed that the cancer development was not suppressed, and the combined cell generated malignant cells, when the nucleus of a non-malignant cell was delivered into a malignant cell's cytoplasm (Figure 3) (Israel and Schaeffer, 1988, Seyfried et al., 2014). Moreover, the delivery of normal mitochondria into cancer cell's cytoplasm resulted in suppression of oncogenic pathways, as well as cancer cell development and proliferation (Elliott et al., 2012, Kaiparettu et al., 2013). In contrast, the combination of normal cytoplasm with malignant mitochondria with pathogenic genetic abnormalities enhanced tumorigenesis (Petros et al.,

2005, Seyfried et al., 2014). Additionally, it seems that the mitochondrial structure and morphology, including lipid and protein content, differ in normal cells vs. cancer cells. In malignant cells, mitochondria represented partial or total lack or breakdown of cristae, a phenomena called cristolysis, as well as matrix swelling. It is suggested that these abnormalities in mitochondrial structure could result in alterations in their functions. Finally, these changes are believed to result in less efficient OXPHOS, resistance to apoptosis, and proton leakage (Arismendi-Morillo, 2009, Seyfried et al., 2014). According to these observations, it is suggested that cancer development is more depended on function and structure of mitochondria than the nuclear genomic abnormalities, that are more involved in directing the malignancy development rather than causing it (Seyfried et al., 2014).

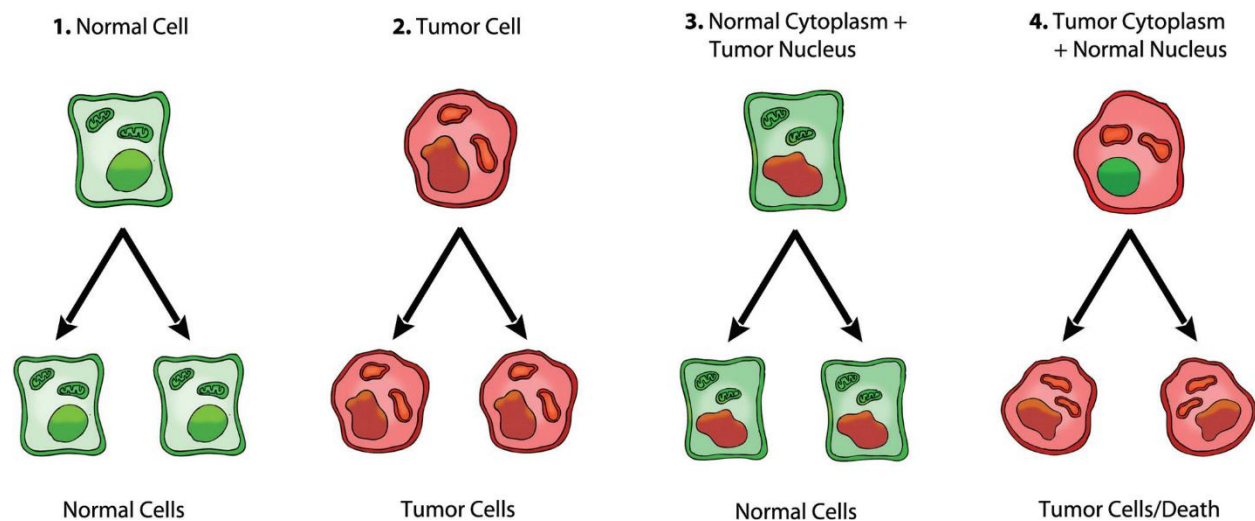


Figure 3. Cytoplasmic Hybrids (Cybrids) Studies.

1. Normal cell generates normal cells. **2.** Tumor cell generates tumor cells. **3.** Transfer of nucleus of a tumor cell to normal cell cytoplasm generates normal cells. **4.** Transfer of nucleus of a normal cell to tumor cell cytoplasm generates tumor cells.

This figure is taken from (Seyfried et al., 2014).

Due to occurrence of functional and structural changes to mitochondria during tumorigenesis, and in accordance with the metabolic view of the cancer's origin, cancer has been suggested to be a mitochondrial metabolic disease. The impairment of mitochondrial respiration, according to this notion, could result in expression of some of the acquired hallmarks of cancer (Figure 4). This hypothesis could be explained through the Lamarckian evolutionary theory. Lamarckian evolutionary theory declares that environmental changes induce inheritable modifications on biological structures to adapt to environment. Differential use or disuse of a biological structure (in this case, Mitochondria) in response to environmental forces results in changes in structure and function to adapt better to the environment. Lamarckian evolution can illustrate the cancer development better than the Darwinian evolution, from the mitochondrial metabolic vantage of the origin of cancer. The mitochondrial adaptation could be transferred into subsequent generations as an acquired feature. By balancing these two evolutionary theories, we could conclude that both changes in mitochondrial functions and genomic mutations could be involved in cancer development (Seyfried et al., 2014).

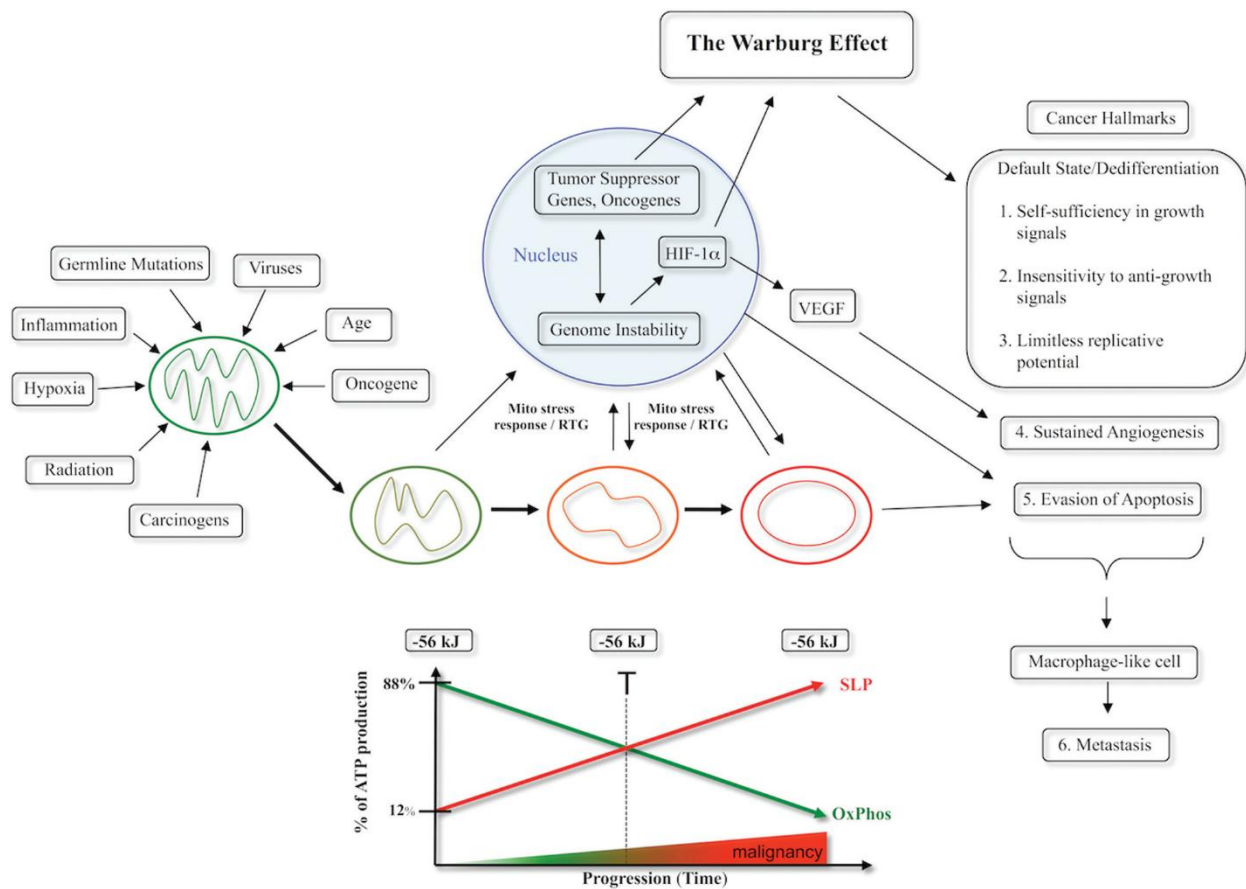


Figure 4. Cancer Originates from Mitochondrial Impairment.

Tumorigenesis occurs in cells with enhanced ATP production via glycolysis (substrate level phosphorylation (SLP)). The metabolic shift from OXPHOS to SLP is directly correlated with the progression of malignancy. According to this view, most hallmarks of cancer are linked directly or indirectly to mitochondrial dysfunction.

This figure is taken from (Seyfried et al., 2014).

Non-Malignant Cells Expressing Warburg Effect:

Not all the cells expressing Warburg effect are cancerous. There are normal cells that are highly proliferative and their bioenergetics is in some ways similar to cancerous cells. For example, proliferating primary lymphocytes utilize >90% of their glucose into aerobic glycolysis (DeBerardinis et al., 2008). Additionally, the highest degree of glycolysis is observed during early

stages of embryogenesis, while it declines slowly during development of an embryo (Warburg, 1956). Warburg effect could be observed in the developing retina of *Xenopus laevis* too. All these observations suggest that aerobic glycolysis is a feature of proliferative metabolism rather than a specific determining hallmark of cancer (Fiske and Vander Heiden, 2012, Seyfried et al., 2014).

The importance of Oncogenes in Glucose Metabolism:

Providing sufficient supplies of vital metabolites, and surviving in limiting oxygen and nutrient conditions, are two of the main problems that proliferative cancer cells face. Many mutations target oncogenes or tumor suppressor genes in order to alter metabolism of cancer cells in the process of developing malignancy (Jones and Thompson, 2009). Here we discuss the most important of these mutations and their effects on metabolism, particularly glucose metabolism, leading to the expression of Warburg effect.

PI3K/Akt/mTOR:

Activation of phosphoinositide 3-kinase (PI3K) by insulin-like growth factor 1 (IGF-1) promotes neuronal survival and immortalization through activation of the serine-threonine kinase Akt (Dudek et al., 1997). Activation of Akt finally leads to activation of mammalian (mechanistic) target of rapamycin, mTOR. The highly conserved PI3K/Akt/mTOR pathway is critically involved in intracellular transmitting signals of growth factors in order to increase metabolism (DeBerardinis et al., 2008). This pathway is also involved in cancer development. The constitutive active version of Akt is mostly observed in tumor cells. Additionally, it could function as the main effector of some of the oncogenic pathways, such as Ras, and BCR-ABL (Elstrom et al., 2004).

An antagonist of PI3K pathway is Phosphatase and TENsin homolog (PTEN). Loss of this tumor suppressor increases the cell division and survival of human cancer cells, as well as alteration of their metabolism (Bost et al., 2016). PI3K and PTEN are involved in cell survival, proliferation, metabolism, and cell-cycle regulation. PTEN degrades phosphatidylinositol (3,4,5)

triphosphate (PIP3), the product of PI3K pathway, to function as the negative feedback of PI3K signaling. Thus, PTEN loss of function mutations could be critically involved in acquiring some of the aforementioned capabilities of cancer cells (Jiang and Liu, 2009, Hanahan and Weinberg, 2011).

The mutations that activate intermediates and effectors of PI3K/Akt/mTOR pathway are among the most common mutations that cause tumorigenesis in humans. These mutations mostly occur on the following genes: PIK3CA, Akt2, PTEN, BCR-ABL, Epidermal Growth Factor Receptor (EGFR), and HER2/neu (ERBB2) (DeBerardinis et al., 2008). Contribution of some of these genes has been found in developing Warburg effect, but their exact role is not clear. For example, in presence to sufficient glucose, activation of PI3K/Akt/mTOR pathway results in a metabolic shift in malignant cells. This metabolic shift leads to elevation of glucose uptake via stimulation of aerobic glycolysis (Elstrom et al., 2004). The expression of genes involved in macromolecule synthesis, such as lipid and proteins, and cell surface transporters of amino acids and glucose could be upregulated by this pathway (DeBerardinis et al., 2008). Akt also participates in positive expression of phosphofructokinase and Hexokinase (particularly Hexokinase II) to force glycolysis (Figure 5) (Jones and Thompson, 2009, Hu et al., 2012). This intriguing direct effect of an oncogene, e.g. Akt, on glycolysis of a cancer cell could have some evolutionary advantages for the development of malignancies. For example, constant glycolysis-mediated preparation of substrates for mitochondrial ETC may inhibit Bax signaling and apoptosis. It also could provide enough supply of substrates for production of amino acids and nucleic acids required for cell growth and division (Elstrom et al., 2004). Additionally, glucose could protect tumor cells against oxidative metabolites such as H₂O₂. Glucose could enter into the pentose phosphate cycle to generate NADPH, which is involved in sustaining the redox balance of the cancer cell (Lord-Fontaine and Averill-Bates, 2002), and increasing its resistance against oxidative environments (Fabregat et al., 1985, Moreira et al., 2016). Lack of glucose could cause cell death by increasing oxidative stress mediated by accumulation of H₂O₂ (Lord-Fontaine and Averill-Bates, 2002).

c-Myc:

In addition to Akt, c-Myc, an early serum response gene, is another oncogene found to alter glucose metabolism of cancer cells (Shim et al., 1997). It belongs to the Myc family of transcription factors containing c-Myc, N-Myc, s-Myc, and L-Myc. They participate in regulation of cell cycle and growth (DeBerardinis et al., 2008). c-Myc is involved in regulation of cell division, differentiation and death, and is commonly expressed in numerous human malignancies. c-Myc trans-activates the lactate dehydrogenase-A (LDH-A) gene (Shim et al., 1997). In normal cells, LDH-A produces lactate from pyruvate anaerobically. Its activity is increased during carcinogenesis. In tumor cells, it is involved in aerobic glycolysis, to convert pyruvate to lactate in presence of oxygen (Shim et al., 1997). Thus, it could be considered as a tumor marker (Shim et al., 1998). c-Myc elevates lactate production by inducing expression of LDH-A. This c-Myc-mediated induction of LDH-A could be critical for neoplastic transformation of human cancer cells, as well as preparing enough supply of metabolites and energy (Shim et al., 1997). In the absence of glucose or inhibition of glycolysis, c-Myc causes apoptosis through an LDH-A mediated manner. This glucose-dependent apoptosis could be inhibited by Bcl-2 activation (Shim et al., 1998). It should be noted that other oncogenes such as HER2/neu and Hypoxia-inducible factor 1 (HIF1) could also activate LDH-A expression (DeBerardinis et al., 2008).

Glucose transporter 1 (GLUT1), phosphofructokinase (PFK), and enolase genes are other direct targets of c-Myc. These genes are critically involved in glucose transport, uptake and glycolysis, and their expression is trans-activated by c-Myc in Rat1 fibroblasts. Thus, c-Myc could be an important regulator of glucose metabolism in cancer cells (Figure 5) (Osthus et al., 2000). In addition to regulation of glucose metabolism, c-Myc plays a role in glutamine metabolism and the pentose phosphate pathway by affecting the expression of various enzymes within these pathways (Hall et al., 2013).

HIF1:

Hypoxia-inducible factor 1 (HIF1) is a transcription factor (DeBerardinis et al., 2008) and a molecular detector of limiting oxygen conditions (hypoxia). HIF1 has two subunits: HIF1 α and

HIF1 β . Under normoxia, the oxygen-sensitive α subunit is degraded via ubiquitination. During cancer development, insufficiency of blood supply to the cancerous tissue results in hypoxia. In the absence of oxygen, the degradation of HIF1 α is inhibited, and it could bind to HIF1 β in the nucleus to create/activate the HIF1 complex. The activated complex stimulates the expression of particular genes such as erythropoietin, vascular endothelial growth factor (VEGF), glucose transporters, heme oxygenase 1, and nitric oxide synthase. These genes critically contribute to glucose metabolism, cell survival and angiogenesis (Kumar et al., 2007). The HIF1 complex could also elevate LDH-A activity, resulting in expression of Warburg effect and enhanced lactate production (Figure 5) (DeBerardinis et al., 2008). Additionally, the synthesis of enzymes involved in glutamine metabolism and the pentose phosphate pathway could be regulated by HIF1 (Hall et al., 2013). It should be noted that both Akt/PI3K and Ras-RAF-MAPK Kinase (MEK)-Extracellular Signal-Regulated Kinase (ERK) pathways can activate HIF1 α expression directly. It has been observed that HIF1 α plays a role in promoting cancer development, as well as suppressing the mismatch repair system, leading to genetic abnormalities (Kumar et al., 2007). Mutations in the von Hippel-Lindau (VHL) causes tumorigenesis by constitutive activation of HIF1 α even in normal oxygen levels. VHL is a tumor suppressor binding to HIF1 α and leading to its ubiquitination and degradation (DeBerardinis et al., 2008).

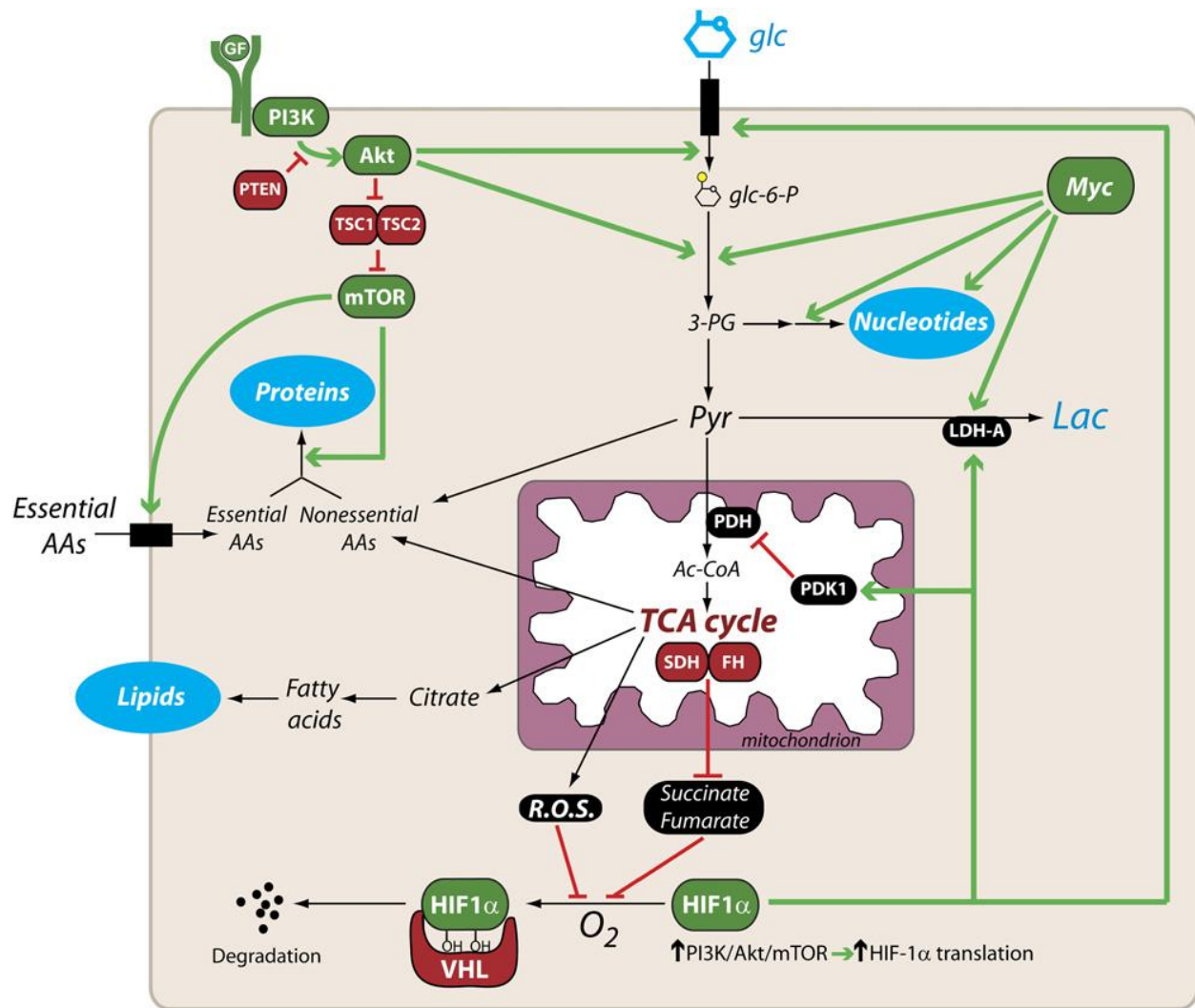


Figure 5. A Signaling Network to Mediate Metabolic Shift in Cancer Cells Leading to Warburg Effect.

This signaling network contains three main pathways: PI3K/Akt/mTOR, Myc, and HIF1. These pathways mostly increase glucose uptake and consumption via glycolysis, but have less effect on mitochondrial respiration. They also stimulate LDH-A to generate more lactate. This increased production of lactate is critically involved in expression of Warburg Effect.

This figure is taken from (DeBerardinis et al., 2008).

Ras:

Ras family of oncogenes is linked to the metabolic reprogramming of cancer cells via disrupting mitochondrial respiration and stimulating glycolysis (Yang et al., 2010, Hu et al., 2012, Neuzil et al., 2012). As mentioned in the “PI3K/Akt/mTOR” section, Ras could activate Akt to trigger its signaling pathway (Elstrom et al., 2004). Mutations of N-, K-, or H-Ras proteins, have been observed in numerous human tumors. Stimulation of Ras signaling cascade increases cAMP levels. This increase activates cAMP-dependent protein kinase A (PKA) in order to phosphorylate and subsequently activate 6-phosphofructo-1-kinase, and finally enhance glycolysis (Yang et al., 2010). Ras could also influence the synthesis of enzymes involved in the pentose phosphate pathway, glutamine metabolism, and glycolysis (Hall et al., 2013). Glutamine metabolism via mitochondria and TCA cycle is required for the cancer cells that are dependent on Ras signaling to proliferate and grow. TCA cycle-mediated catabolism of glutamine is a vital source for production of ATP, NADPH, and biomolecules such as lipids, nucleotides and amino acids (Weinberg et al., 2010).

K-Ras^{G12BV} mutation is observed in numerous human tumors. This mutation results in constitutive activity of K-Ras, particularly in pancreatic cancer. Constitutive activation of K-Ras rapidly alters the function of mitochondria in several ways. It decreases the expression of ETC complex 1 subunits (e.g. NDUFA2, NDUFA4, NDUFA5, NDUFA11, NDUFA12, NDUFA13, NDUFB4, NDUFB6, & NDUFB7), catalase, and superoxide dismutase 2 (SOD2). It increases the expression of ETC complex 2, as well as stimulating the aerobic glycolysis, ROS production, and upregulation of antioxidant glutathione (GSH) expression (Hu et al., 2012, Neuzil et al., 2012). In addition to K-Ras^{G12BV}, the constitutive activating H-Ras^{Q61L} mutant inhibits mitochondrial respiration and decreases oxygen consumption. Instead, it increases glycolysis, leading to independence of bioenergetics from OXPHOS. It should be noted that cellular transformation of Ras requires the association of Signal Transducer and Activator of Transcription 3 (STAT3) in mitochondria. The transformed cells with H-Ras^{Q61L} are very sensitive to glucose deficiency, and could not use glutamate or pyruvate alternatively (Yang et al., 2010).

BRAF:

Another effector of Ras signaling, BRAF, is highly mutated in some of the cancers, particularly melanomas and thyroid cancers. BRAF belongs to RAF family of serine/threonine kinase including 3 members called ARAF, BRAF, and CRAF. They function as intermediate intracellular signaling kinases in Mitogen-Activated Protein Kinase (MAPK) pathway regulating expression of genes involved in cell survival and proliferation (Kumar et al., 2007). The Ras-BRAF-MEK-ERK pathway is involved in controlling hallmarks of cancer, such as angiogenesis, cell survival, cell division, invasion of cancer cells, and metabolic features leading to the Warburg effect (Kumar et al., 2007, Hall et al., 2013).

The constitutive active BRAF^{V600E} mutant results in positive regulation of glycolytic genes and subsequently elevation of glycolysis, and reduction of oxygen consumption (Hall et al., 2013). This mutant inhibits apoptosis of melanoma cells exposed to low levels of oxygen (Kumar et al., 2007). It seems that the BRAF-mediated activation of MEK & ERK could be involved in glucose metabolism via stimulation of HIF1 α or c-Myc. The Ras-RAF-MEK-ERK signaling cascade could directly upregulate both HIF1 α and c-Myc expressions (Yeh et al., 2004, Kumar et al., 2007, Hall et al., 2013).

TP53:

Tumor suppressor protein p53 (TP53) is a transcription factor activated in response to cellular stressful conditions. These conditions include low levels of oxygen, oxidative environment, metabolic insults, and genomic instabilities. In response to these conditions, TP53 stimulates transcription of genes critically involved in apoptosis, senescence, and inhibition of cell division, resulting in adaptation of the cell to new conditions. TP53 could be activated by mTOR and the Adenosine monophosphate (AMP)-activated protein kinase (AMPK). AMPK is a heterotrimeric protein kinase complex detecting the cellular AMP:ATP ratio. It is critically involved in controlling the balance of cellular bioenergetics. AMPK activates TP53 via phosphorylation of its Serine 15. The activated TP53 positively regulates expression of OXPHOS genes. Finally, its activation directs to a shift in bioenergetic homeostasis toward obtaining ATP

from OXPHOS (Jones and Thompson, 2009). As the mutations of TP53 are frequently observed in tumors, it has been suggested that TP53 could be one of the genes responsible for expressing Warburg effect in cancer cells (Matoba et al., 2006). TP53 increases the biosynthesis of Cytochrome C Oxidase subunit 2 (SCO2) of complex IV of ETC and TP53-induced glycolysis and apoptosis regulator (TIGAR). TIGAR inhibits glycolysis by decreasing fructose-2,6-bisphosphate. Fructose-2,6-bisphosphate is a stimulator of PFK1, a regulatory enzyme in glycolysis. SCO2 is required for assembly and regulation of complex IV which is the major consumer of oxygen in the cell. Deficiency of TP53 in cells results in disruption of SCO2 and subsequently impairment of ETC, leading to a metabolic shift from OXPHOS to glycolysis to obtain ATP (Bensaad et al., 2006, Matoba et al., 2006, Jones and Thompson, 2009).

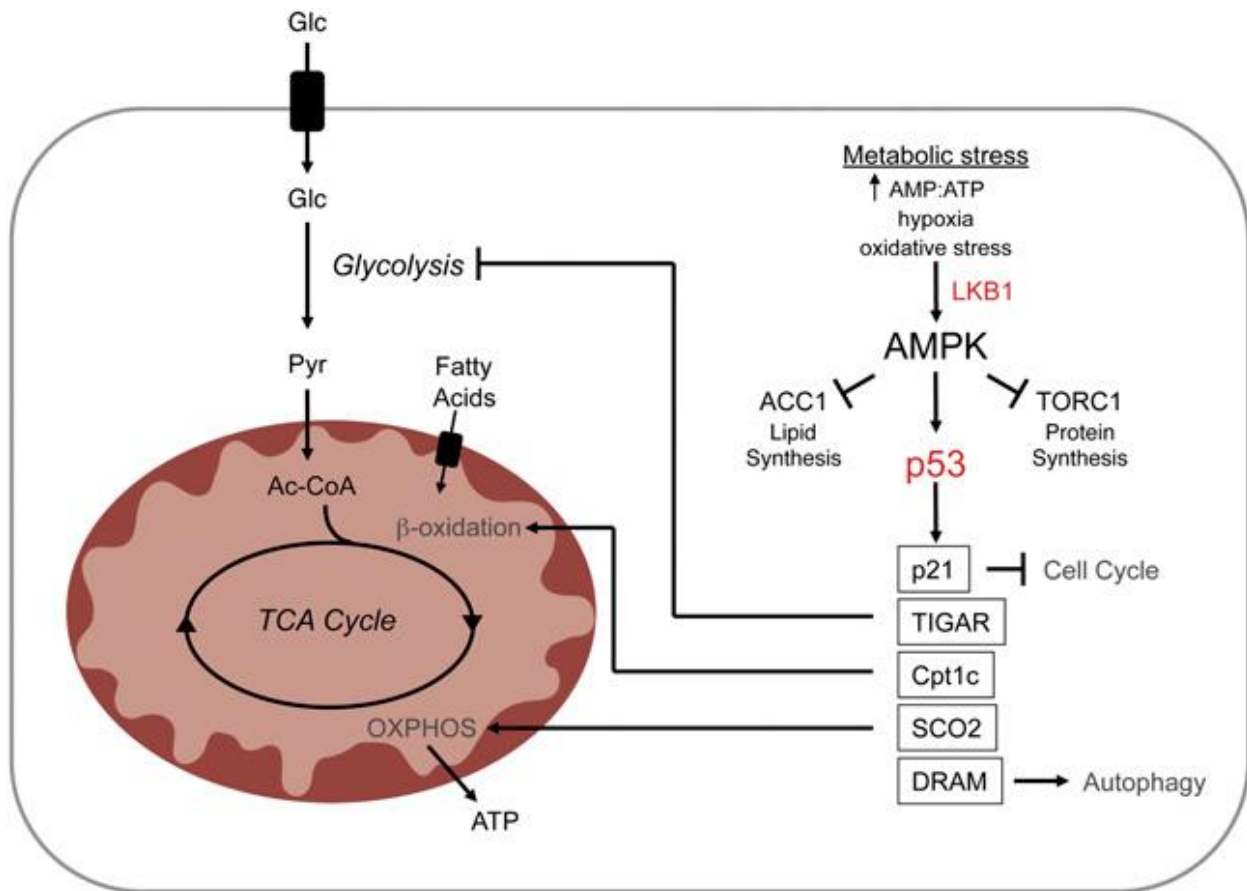


Figure 6. The AMPK-TP53 Pathway.

Activation of AMPK-TP53 pathway triggers a subset of effectors activation. Activation of p21 and TIGAR inhibits cell cycle and glycolysis, respectively. Activation of DRAM, SCO2, and Cpt1c stimulates autophagy, OXPHOS, and β-oxidation of fatty acids, respectively. Inhibition of this pathway, reverses all these effects, results in expression of Warburg Effect.

TORC1: TOR Complex 1, ACC1: Acetyl-CoA Carboxylase 1, Cpt1c: Carnitine palmitoyltransferase 1c, DRAM: Damage-Regulated Autophagy Modulator,

This figure is taken from (Jones and Thompson, 2009).

Phylogenetic Profiling:

First proposed by Pellegrini *et al.* in 1999, phylogenetic profiling is a method based on the evolutionary hypothesis that states proteins that function together prefer to evolve together. According to this notion, evolution prefers to keep or eliminate functionally related proteins in a new organism in a coordinated manner (Pellegrini *et al.*, 1999). The gene elimination, gain, speciation, and lateral transfer are processes involved in the evolutionary journey of a gene/protein through different species (Basu *et al.*, 2011). The evolution of functionally related proteins creates a regular pattern of presence or absence of their homologs in a subset of genomes. This correlated pattern is called phylogenetic profiling (Pellegrini *et al.*, 1999). Thus, phylogenetic profiling shows the appearance or disappearance of a gene/protein across genomes of multiple species. This method is used for finding functional/evolutionary associations of genes/proteins, predicting their roles, and discovering complex organization of molecular systems and pathways (Cui *et al.*, 2011, Kotaru *et al.*, 2013).

The amount of data obtained from sequenced genomes and proteins are rapidly increasing. Additionally, experimental procedures are limited in identifying function of proteins and genes. Tools/methods for predicting functions and finding relationships of genes/proteins are required for saving time and expenses. Phylogenetic profiling is one of these methods. Its simplest form for a gene/protein is a binary vector which has a length of the number of analyzing genomes, representing 1 or 0 for each of them, if the homologue of the referring gene/protein is present or absent in the target genome, respectively (Kotaru *et al.*, 2013). Then, it is possible to identify the co-evolution or correlation of various groups of genes/proteins, such as enzymes involved in a particular metabolic pathway or some subunits of protein complexes, by comparing their phylogenetic profiles to predict the function of unknown genes/proteins (Figure 7) (Pellegrini, 2012). The predictive results of phylogenetic profiling concerning the functions of a number of unknown proteins have been confirmed by experimental results (Kensche *et al.*, 2008). For example, phylogenetic profiling was used in a study to predict the functions of mitochondrial proteins, and its results were compared with another method, comparative physiology. The assumption of this study was based on the endosymbiotic theory. According to this theory, the origin of mitochondria returns to a free-living alpha-proteobacteria transformed

to an organelle during the intracellular symbiosis with the origin of eukaryotic cells. Then, the new organelle has lost or gained genes through its evolution from yeast to human and generated an evolutionary memory or footprint. This evolutionary footprint could be demonstrated and traced by phylogenetic profiling (Cheng and Perocchi, 2015).

Four factors are involved in affecting the predicted results of the method: 1) Selection of Target Taxa: Generally, increasing the number of genomes results in more accurate results. Elimination of redundant species will improve the results. 2) Identification of Homologs of a Reference Protein: Applying sequence similarities methods, such as Basic Local Alignment Search Tool (BLAST), could critically affect the predicted results. These methods find the best group of homologs for the reference gene/protein. 3) Illustration of Homolog Distribution: Presence or absence of a homolog in a genome could be demonstrated by binary characters (i.e. integers of 1 or 0). In other methods, the sequence similarity results (e.g. BLAST) could show the presence or absence of the homolog in a fuzzy scale between 0 & 1. 4) Scoring Method of Phylogenetic Profiles: Selecting an appropriate method to score and compare phylogenetic profiles is critical for receiving the best result. There are various approaches calculating phylogenetic profiles. Scoring methods can calculate simple distance measures (e.g. Jaccard coefficients), or statistical parameters (e.g. Pearson's correlation coefficient) (Basu et al., 2011). These methods stand on this assumption that presence/absence of a gene in a particular genome is independent of the presence/absence of its homologue in another genome. Due to this independence, they are called "Naïve Co-Occurrence Profiling Methods" (Kensche et al., 2008). There are also other methods (e.g. Kernel based Method) considering evolutionary processes such as gene elimination and speciation from the species to their ancestors, as well as the aforementioned parameters (Basu et al., 2011). They are called "Model-based Co-Occurrence Methods". Unlike Naïve measures, model-based methods consider evolutionary dependency of occurrence values of a gene in a set of genomes (Kensche et al., 2008).

Similar to other bioinformatics techniques, there are some limitations and disadvantages to the application of phylogenetic profiling. Simplification of complexity of biological pathways and mechanisms into numbers results in loss of information. This loss of information limits the yield and coverage of the phylogenetic profiling method. For example, accurate functional

predictions can be obtained if a high score threshold is applied and only the most similar phylogenetic profiles are selected. If the score threshold decreases, more false positive predictions will be obtained. The other limitation of this technique originates from the type of relationships between genes/proteins. Most of co-occurrence methods (e.g. Hamming distance method) could investigate symmetric and binary relationships. In symmetric relationships, there is no difference between association of *protein 1* with *protein 2* and vice versa. It means that the measured values of association of *protein 1* with *protein 2*, and association of *protein 2* with *protein 1*, are equal. This inability of the methods to differentiate between and association of *protein 1* with *protein 2*, and vice versa, leads to another disadvantage of the technique (Kensche et al., 2008). This disadvantage of the method results in inaccurate functional predictions for most proteins involved in a metabolic pathway. Except for a few core proteins conserved evolutionarily, most of the proteins of a metabolic pathway have irregular patterns and distributions in their phylogenetic profiling. These irregularities in distributions of most metabolic proteins could be a representative of their adaptation to the circumstances of species or their environments (Muley and Ranjan, 2013). Thus, asymmetric relationships are required to be considered. Unlike symmetric relationships, asymmetric relationships differentiate between association of two genes/proteins in both directions. Asymmetric relationships occur during the evolution of a complex or pathway, especially when some accessory subunits are gained or lost from an evolutionarily conserved pathway. Asymmetric relationships can be calculated by model-based methods, such as maximum likelihood method (Kensche et al., 2008).

The other limitation of phylogenetic profiling relates to the number and phylum of sequenced genomes. Generally, increasing the number of genomes improves the results of the functional prediction. In another words, it increases the resolution of phylogenetic profiles (Kensche et al., 2008). Due to this limitation, Lin et al. (2013) claimed that this technique is not suitable for analyzing eukaryotic genes/proteins, and works properly only on prokaryotes. The number of completely sequenced genomes of prokaryotes are much more than the eukaryotes, and the organization of their genomes are different with each other. Due to these reasons, the application of prokaryotic genomes for evaluating eukaryotic phylogenetic profiling is limited (Lin et al., 2013). Another reason suggested for this limitation is the modular evolution of eukaryotic

proteins, since some of them are the results of parasitism and endosymbiosis (Simonsen et al., 2012). The organization of biological systems are modular. Each system includes sub-systems restricted spatially and biochemically, performing their specialized functions. A module is a group of proteins that interact with each other and are limited to a subcellular region, or a group of enzymes involved in a metabolic pathway (Dey and Meyer, 2015). An evolutionary module is a set of proteins whose phylogenetic profiles are more similar to each other in comparison with other proteins outside of the module. In another words, the proteins of a module are co-evolved with each other (Kensche et al., 2008). It seems that evolution works mostly on small sub-modules rather than large modules (Dey et al., 2015). One possible way for solving the complexity of modular evolution could be selecting the most suitable reference genomes to reach to the highest accuracy in protein function prediction, in a pathway-dependent manner. It has been suggested that selecting reference genomes from the archaea, bacteria, and eukaryotes could be the best approach for highly conserved pathways. For specialized networks, selecting reference genomes from close taxa is highly suggested for reaching to best results (Simonsen et al., 2012). Another suggested solution is categorization of modular evolution of proteins (Orthogroups). By considering this concept in the algorithm of the method, signaling and transcriptional cascades could be differentiated from protein complexes or metabolic pathways (Dey et al., 2015).

To accurately predict functions of unknown proteins, some essential criteria need to be considered, including: selecting the method, scoring threshold, referenced genomes, and evolutionary modular groups of proteins. Although phylogenetic profiling predictions could be highly accurate, they are required to be confirmed by experimental evidence (Kensche et al., 2008).

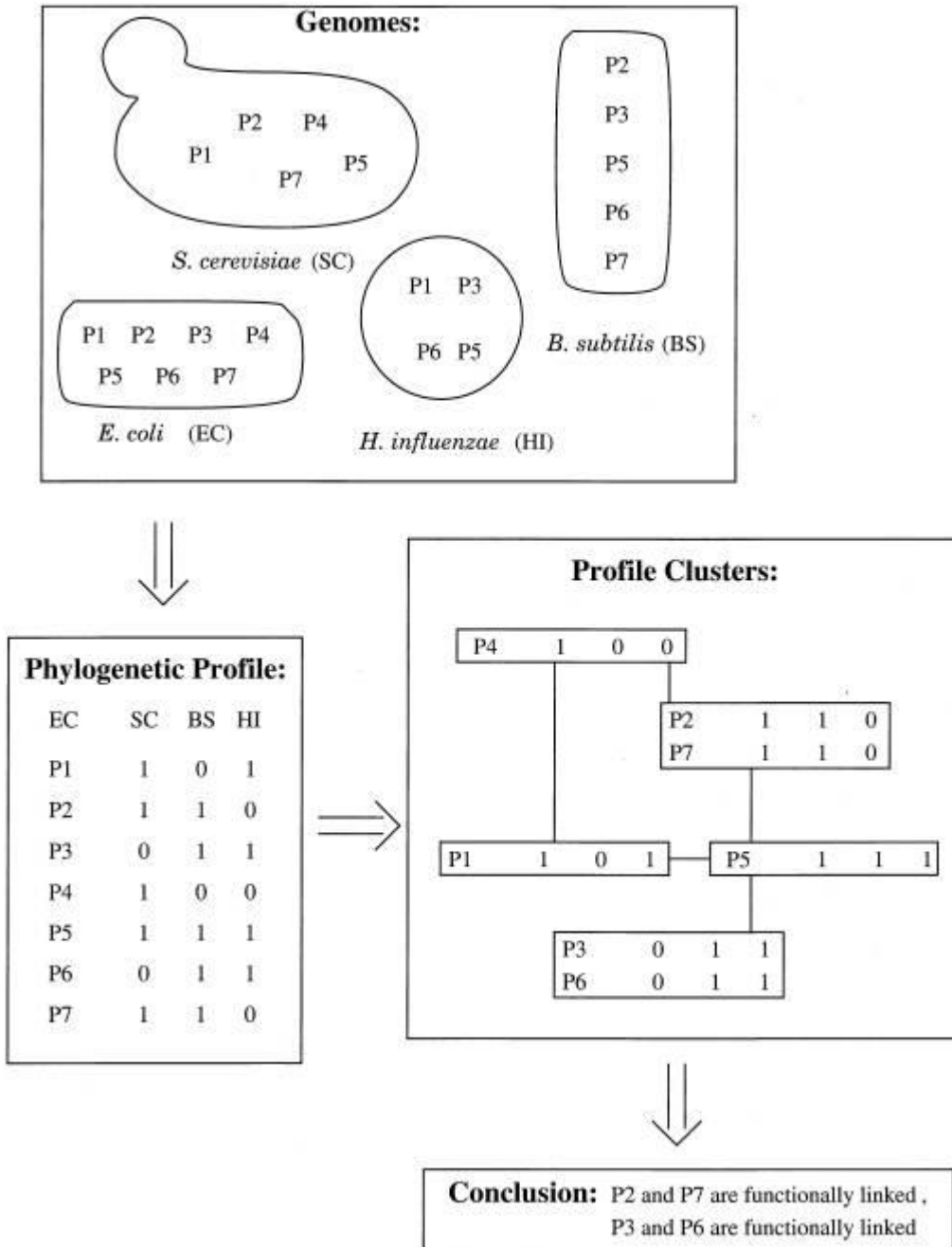


Figure 7. A Schematic Illustration of Phylogenetic Profiling Method of Binary Vectors.

This method shows analysis of 7 protein phylogenetic profiles in 4 species, and prediction of their functional linkages.

This figure is taken from (Pellegrini et al., 1999).

Overall Hypothesis and Objectives:

Since experimental studies have shown that the metabolic shifts in glycolysis, TCA cycle, and OXPHOS, involved in bioenergetics and biosynthesis of cancer cells, are correlated with some frequently-mutated oncogenes (the Glycolytic-TCA-OXPHOS-Correlated Oncogenes or GTO-Oncogenes), we hypothesized that the GTO-Oncogenes and genes of glycolysis, TCA cycle, and OXPHOS pathways could have evolved to function together. The overall objective of this study was to determine whether there is a correlation between the phylogenetic profiles of the GTO-Oncogenes and the genes involved in glycolysis, TCA cycle and OXPHOS. Another aim of the study was to predict the specific potential roles of GTO-Oncogenes in modulation of the metabolic pathways, and to find other oncogenes that may be involved in the metabolic reprogramming of the cancer cells.

Methods and Materials:

Selection of Oncogenes:

All of the “Significantly mutated genes” and “Additional significantly mutated genes found under restricted hypothesis testing (RHT)” identified in 21 types of tumor were extracted from the Supplementary Table 3 [File: 1. Supplementary Table 3 (Lawrence et al., 2014).xlsx] of (Lawrence et al., 2014). A total number of 233 oncogenes were obtained and imported into an excel file for statistical analysis [File: 2. List of Oncogenes.xlsx]. The number of repetitions was calculated for each of 233 oncogenes. All of the oncogenes with the repetition number of ≥ 2 were primarily determined. A total number of 98 genes were obtained and selected in this step. The final selection of oncogenes was performed by removing 11 of “GTO-Oncogenes” from the list. So, a total number of 87 oncogenes were selected [File: 2. List of Oncogenes.xlsx – Sheet 2].

Selection of GTO-Oncogenes:

According to our literature review (summarized in the “Introduction” section), the following oncogenes were selected as GTO-Correlated candidates: TP53, PIK3CA, PTEN, KRAS,

NRAS, HRAS, BRAF, HER2/neu (ERBB2), c-Myc, EGFR, Akt1, Akt2, Akt3, ABL (bcr/abl), HIF-1 α , HIF-1 β , LDH-A, and mTOR. After comparing the phylogenetic profiles of these oncogenes, the oncogenes with low correlation (<50) were eliminated from the list of candidates. The chromosomal position of these oncogenes were retrieved from “Kyoto Encyclopedia of Genes and Genomes” (KEGG) database (Kanehisa et al., 2016) available at <http://www.genome.jp/kegg/kegg2.html> (retrieved on 08/01/2016). The mitochondrial localization of them was investigated via the Human MitoCarta2.0 database (Calvo et al., 2016) available at <https://www.broadinstitute.org/scientific-community/science/programs/metabolic-disease-program/publications/mitocarta/mitocarta-in-0> (retrieved on 08/01/2016).

Selection of Genes of Glucose Metabolism:

The *hsa* Gene IDs and annotations of all human genes involved in glycolysis/gluconeogenesis/ pyruvate oxidation pathway (67 genes), TCA cycle (30 genes), and oxidative phosphorylation (109 genes) were obtained from KEGG database (Kanehisa et al., 2016). Seven of these genes were shared with glycolysis/gluconeogenesis/ pyruvate oxidation pathway and TCA cycle: DLAT, DLD, PCK1, PCK2, PDHA1, PDHA2, & PDHB. Another four genes of TCA cycle were shared with Succinate Dehydrogenase Complex (Complex 2) of ETC: SDHA, SDHB, SDHC, & SDHD. Thus, a total number of 195 genes were incorporated into a single excel file [File: 3. Genes of Glucose Metabolism]. The OXPHOS genes were categorized according to their respective mitochondrial complexes, except for 5 genes. These genes did not belong to any of the complexes: ATP12A, ATP4A, ATP4B, LHPP, PPA1, & PPA2. Random gene sets were used as negative control.

Construction of Phylogenetic Profiles:

Construction of phylogenetic profiles was performed using a tool developed in our laboratory PrePhyloPro Software version 0.1.7 available at

<http://prephylopro.org/phyloprofile/>> (retrieved on 08/01/2016). We imported genes of different categories (e.g. 89 oncogenes, 17 GTO-Related candidates, and 195 genes of glucose metabolism) into a csv file along with their Gene IDs, annotations, and Gene Colors. The *hsa* Gene IDs and annotations of all aforementioned oncogenes for human species (*Homo sapiens*) were obtained from the “Input gene list”. After preparation of our inputs list, we set the algorithm parameters of the PrePhyloPro Software as follows: 1) Linkages number: the value of 100 was selected. 2) BLAST Eval: the minimum number of 0.0001 was selected. 3) Organism: *Homo sapiens* (Human) was selected for this parameter. Then, the gene list was uploaded and analyzed by the software. The phylogenetic profiling was performed based on 972 taxa (82 archaea, 614 bacterial, & 276 eukaryotic species including chloroplast & mitochondrial proteins). The output files of the software are “Correlation Matrix” table and plot, profile plot, and predicted linkages. The software uses the Pearson correlation coefficient and Jaccard similarity measures (Choi et al., 2010) for calculation of correlation/similarity between phylogenetic profiles.

Results:

To compare the phylogenetic profiling of glucose metabolism genes, GTO-Oncogenes, and other oncogenes, we performed several analyses. First we selected the following oncogenes, based on literature review, to construct and compare their phylogenetic profiles: TP53, PIK3CA, PTEN, KRAS, NRAS, HRAS, BRAF, HER2/neu (ERBB2), c-Myc, EGFR, Akt1, Akt2, Akt3, ABL (bcr/abl), HIF-1 α (HIF1 α), HIF-1 β (HIF1 β), LDH-A, and mTOR. Then, the phylogenetic profile of these oncogenes were constructed and analyzed using the PrePhyloPro Software version 0.1.7 (Linkages Number = 100, Blast Eval = 0.0001, *Homo sapiens*) [File: 1. GTO-Oncogenes.tar.gz]. We adjusted the Blast Eval to the smallest value in the software (0.0001) to increase the accuracy of the functional predictions. By applying a high score threshold (or small Blast Eval) only the most similar homologs are selected, improving the accuracy of our predicted results (Kensche et al., 2008).

A comparison of the correlation coefficients of the phylogenetic profiles of the oncogene pairs revealed that LDH-A had the lowest correlation (i.e. ~0) with the other input genes. Thus, it

was excluded from the list of inputs. Then, we grouped the oncogenes based on the similarity/correlation of their phylogenetic profiles. According to their occurrence patterns (presence/absence) in different kingdoms & phyla, these oncogenes were divided into 3 subgroups (Figure 8): 1) Oncogenes mainly found in Animals (HIF-1 β , HIF-1 α , c-Myc, & TP53) with correlations of 0.50-0.70. 2) Oncogenes mostly found in Eukaryotes (PTEN, PIK3CA, mTOR, N-Ras, K-Ras, & H-Ras) with correlations of 0.50-1.00. 3) Oncogenes found in both Eukaryotes and Prokaryotes (HER2/neu, EGFR, ABL, BRAF, Akt1, Akt2, & Akt3). These seem to be highly conserved in all kingdoms of life. The correlations of these genes are 0.80-1.00.

Based on the comparison of correlation scores, it was obvious that the phylogenetic profiles of Subgroup 3 oncogenes have the most similarity with each other. Although some of them were from the same gene family (e.g. Akt 1, 2, and 3), the overall high degree of conservation from prokaryotes (particularly eubacteria) to animals could represent a conserved biochemical/signaling pathway. On the other hand, the oncogenes of other subgroups, especially Subgroup 1, were not highly correlated with each other. One of the objectives of this study was to compare phylogenetic profiles of the most relevant GTO-Oncogenes with metabolism genes of glycolysis, TCA cycle, and OXPHOS. Because the phylogenetic profiles of some of the GTO-Oncogenes were not highly correlated with each other, we performed additional analysis to narrow down the list of GTO-Oncogenes. For this reason, the average mutation rate and the number of repetition of 11 of the above oncogenes were calculated, based on the data from (Lawrence et al., 2014). The others were not found in this study. Also, the gene position and mitochondrial localization of all of the 17 GTO-Oncogenes were investigated [File: 2. Additional Analysis of GTO-Oncogenes.xlsx]. Based on our results, none of these oncogenes were considered mitochondrial proteins. The most repeated & mutated oncogenes were TP53 (Subgroup 1, #Repeat = 18, Average Mutation Rate = 34.9%), PTEN (Subgroup 2, #Repeat = 9, Average Mutation Rate = 14.87%), K-Ras (Subgroup 2, #Repeat = 9, Average Mutation Rate = 14.23%), N-Ras (Subgroup 2, #Repeat = 8, Average Mutation Rate = 8.15%), PIK3CA (Subgroup 2, #Repeat = 8, Average Mutation Rate = 14.13%), BRAF (Subgroup 3, #Repeat = 6, Average Mutation Rate = 15.55%), HER2/neu (Subgroup 3, #Repeat = 6, Average Mutation Rate = 2.72%), EGFR (Subgroup 3, #Repeat = 4, Average Mutation Rate = 10.63%), H-Ras (Subgroup 2, #Repeat = 3, Average

Mutation Rate = 3.93%), mTOR (Subgroup 2, #Repeat = 2, Average Mutation Rate = 8.45%), and Akt1 (Subgroup3, #Repeat = 2, Average Mutation Rate = 2.65%). For better visualization of the mutation rate, we decided to classify them. By selecting 10.0% as the threshold for mutation rate, we divided our GTO-Oncogenes into 2 classes: High & Low mutated GTO-Oncogenes (Table 1).

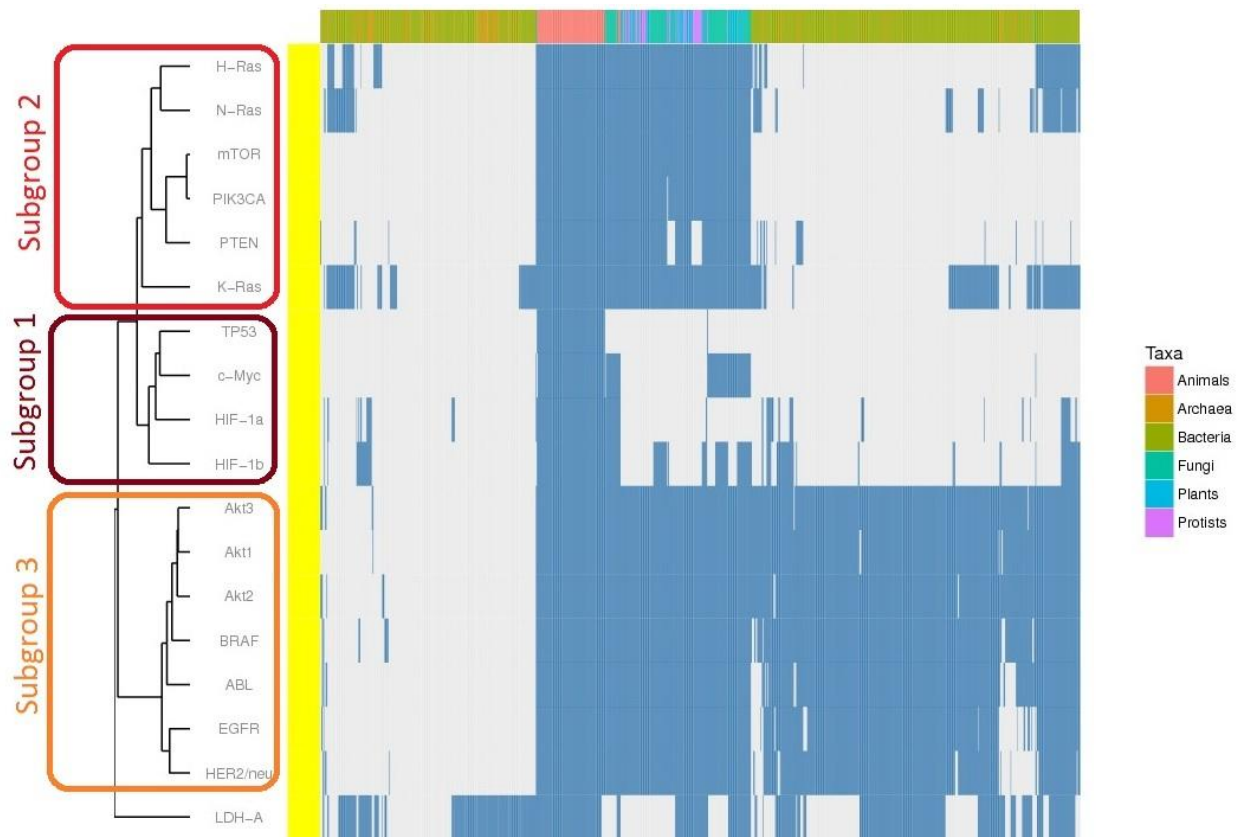


Figure 8. Categorization of Oncogenes involved in Metabolic Shift of Glycolysis, TCA Cycle and OXPHOS in Cancer Cells (GTO-Oncogenes).

According to presence or absence of each oncogene in different kingdoms and phyla, they were categorized into 3 subgroups. Subgroups 1, 2, and 3 were found in “Animals”, “Eukaryotes”, and “Eukaryotes & Prokaryotes”, respectively. Due to very low correlation of LDH-A with other inputs, it was excluded from analysis.

	Subgroup 1 (found in Animals)	Subgroup 2 (found in Eukaryotes)	Subgroup 3 (found in Prokaryotes & Eukaryotes)
High Mutation Rate ($\geq 10.0\%$)	TP53	PIK3CA, PTEN, K-Ras	BRAF, EGFR
Low Mutation Rate ($< 10.0\%$)	c-Myc, HIF1 α , HIF1 β	mTOR, N-Ras, H-Ras	ABL, Akt1, Akt2, Akt3, HER2/neu

Table 1. Classification of GTO-Oncogenes based on Their Mutation Rate.

By selecting 10.0% as the threshold for mutation rate, our 17 oncogenes were divided into 2 classes: High and Low mutated oncogenes. By combination of this classification with their categorization according to their occurrence pattern, 6 groups were obtained.

To find the best correlated GTO-Oncogenes, we selected the oncogenes with high mutation rate from all the 3 subgroups and compared their phylogenetic profiling correlations (File: 3. Correlation Matrix of High Mutated Oncogenes.xlsx). As TP53 was the only member of Subgroup 1 with high mutation rate, there was no need to compare it with the other members of its subgroup. From comparison of high mutated members of Subgroup 2, K-Ras had the lowest correlation with PIK3CA and PTEN, with the value of 0.61 & 0.54, respectively. Thus, it was eliminated from our final list. The high mutated members of Subgroup 3, i.e. BRAF and EGFR, had the correlation of 0.87 with each other. Therefore, both of them were selected as the representatives of their subgroup. Thus, TP53 was selected as the representative of Subgroup 1, PIK3CA and PTEN as the representatives of Subgroup 2, and BRAF and EGFR as the representatives of Subgroup 3.

To determine whether there was a correlation between the GTO-Oncogenes and the genes of glucose metabolism, we compared the phylogenetic profiles of the two gene sets.

Analysis of the phylogenetic profiles would provide novel insight into the potential interactions of these oncogenes with genes of different stages of glucose metabolism in cancer cells. Although none of these oncogenes was a mitochondrial protein [File: 2. Additional Analysis of GTO-Oncogenes.xlsx], some of them could translocate to mitochondria under certain conditions. Under stress, the wild-type TP53 moves to mitochondrial outer membrane to trigger apoptosis. It also could enter mitochondrial matrix to protect its genome (Vaseva and Moll, 2013). PTEN is another representative of GTO-Oncogenes which its localization into mitochondria is critical for mitochondrial-dependent apoptosis (Zhu et al., 2006). Interestingly, no mitochondrial localization is found for wild-type BRAF, while its activating mutants such as BRAF (G469A), BRAF (V600E), BRAF (V600K), and BRAF (V600D), could translocate to mitochondria to exert their anti-apoptotic effects (Lee et al., 2011). In the presence of EGF, EGFR could translocate to mitochondria and bind to COX2 (SCO2). Subsequently, EGFR could phosphorylate COX2 to modulate mitochondrial functions (Demory et al., 2009). All these studies showed that we could expect to find phylogenetic correlations between our oncogenic representatives and glucose metabolic genes in the cytoplasm or the genes of TCA cycles and OXPHOS, found in mitochondria.

The Phylogenetic Profile of TP53 is Highly Correlated with Complexes 1, 4 and 5 of Mitochondrial ETC.

For determining which components of glucose metabolic pathways were highly correlated with the TP53 phylogenetically, a list of glucose metabolic genes was required. For this reason, 195 genes involved in glucose metabolism, glycolysis/gluconeogenesis/pyruvate oxidation pathway + TCA cycle + oxidative phosphorylation, were retrieved from KEGG database (Kanehisa et al., 2016). The phylogenetic profiles of all these genes as well as TP53 were prepared and analyzed by the PrePhyloPro Software version 0.1.7 (Linkages Number = 100, Blast Eval = 0.0001, Homo sapiens) [File: 4. TP53 & Genes of Glucose Metabolism.tar.gz]. For better comparison of the phylogenetic profiles, we selected the cut off value of 0.85. 17 glucose metabolic genes had similarity correlation of ≥ 0.85 with TP53: ATP4B, ATP5L, ATP5J, ATP5J2, COX4I2, COX6C, COX7A2L, COX7A2, COX7A1, NDUFB11, NDUFB9, NDUFA7, NDUFB8, NDUFB10,

NDUFB1, NDUFB5, & G6PC2. Three of these genes were from ATP Synthase Complex. Five were from Cytochrome C Oxidase Complex (Complex IV). Seven were from Complex I of ETC. One of the oxidative phosphorylation genes, ATP4B, had the highest correlation with TP53 with the value of 0.98. Interestingly, it did not belong to any of the ETC complexes. The phylogenetic profile of one of the genes from glycolysis/gluconeogenesis/pyruvate oxidation pathway, i.e. G6PC2, also displayed a high correlation with the TP53 gene.

It seemed that TP53 was highly correlated with a select group of OXPHOS components (Complexes I, IV, & ATP Synthase), rather than the components of TCA cycle and glycolysis/gluconeogenesis/pyruvate oxidation pathway.

The Phylogenetic Profiles of only PIK3CA, but not PTEN, are Highly Correlated with Complexes 1, 3, and 4 of Mitochondrial ETC, as well as Hexokinases of Glycolysis Pathway.

Similar to analysis of TP53, we identified which genes of glucose metabolism were highly correlated with the representatives of Subgroup 2, PIK3CA and PTEN. The phylogenetic profiles of all the aforementioned 195 genes involved in glucose metabolism as well as PIK3CA and PTEN were constructed and analyzed by the PrePhyloPro Software version 0.1.7 (Linkages Number = 100, Blast Evaluated = 0.0001, Homo sapiens) [File: 5. Subgroup 2 Representatives & Genes of Glucose Metabolism.tar.gz]. The threshold of 0.85 was selected for phylogenetic correlation coefficients. PTEN did not have a correlation coefficient of ≥ 0.85 with the aforementioned metabolic genes. At this threshold, PIK3CA had correlation with 13 metabolic genes: COX6B1, COX6B2, COX5B, COX17, UQCRB, NDUFA8, NDUFA13, NDUFA6, HK1, HK2, HK3, GCK (HK4), & HKDC1. From these 13 genes, four genes were from Cytochrome C Oxidase Complex (Complex IV). Three genes belonged to Complex I. One gene was from Cytochrome bc1 Complex (Complex III), and five genes (Hexokinases) were glycolytic enzymes.

Based on these results, only PIK3CA was highly correlated with some components of OXPHOS (Complexes I, III & IV) as well as Hexokinases. No significant correlation was observed between the phylogenetic profiles of this GTO-Oncogene and those of the other components of glycolysis/gluconeogenesis, TCA cycle or ETC.

There's no Significant Correlation between the Phylogenetic Profiles of BRAF and EGFR with Components of Glycolysis/Gluconeogenesis, TCA, and OXPHOS.

In the third round of our analysis, we investigated the phylogenetic correlation of the representatives of Subgroup 3 of GTO-Oncogenes, BRAF and EGFR, with the aforementioned 195 metabolic genes. As these subgroup oncogenes were found in prokaryotes and eukaryotes, it would be predictable that they likely would have high correlation with highly conserved pathways or complexes. Thus, their phylogenetic profiles were generated and analyzed by the PrePhyloPro Software version 0.1.7 (Linkages Number = 100, Blast Evalule = 0.0001, Homo sapiens) [File: 6. Subgroup 3 Representatives & Genes of Glucose Metabolism.tar.gz]. Interestingly, at threshold of 0.85, no correlation was observed between phylogenetic profiles of BRAF or EGFR and glucose metabolic genes.

According to these results, we could observe that there was no significant correlation between phylogenetic profiles of BRAF and EGFR with all the 195 genes of glucose metabolism. It was against our prediction stating that these oncogenes would likely be correlated with highly conserved pathways. Perhaps this observation showed that the Subgroup 3 representatives did not gain any direct interactions with components of cellular bioenergetics during evolution. Most of the genes of glucose metabolism found in both prokaryotes and eukaryotes did not show any significant correlation with each other. As is shown in Figure 9, most of these metabolic genes had low correlation with each other, demonstrated by blue (correlation coefficient 0.0-0.25) and gray spots (correlation coefficient <0.0). It seems that they preferred to form cluster with each other in small numbers and with distinct borders (Figure 9). Because of that, these glucose metabolic genes demonstrated high correlation within their clusters, while lacking significant correlation with other clusters.



Figure 9. Formation of Clusters in Glucose Metabolic Genes found in Prokaryotes and Eukaryotes.

The phylogenetic profiles of glucose metabolic genes found in prokaryotes and eukaryotes had no significant correlation with each other, except within clusters. Because of that, we can see red spot clusters (correlation coefficient >0.75) at the edge a blue-gray context (correlation coefficient <0.25). This observation could show independent evolution of different pathways and complexes.

Phylogenetic Profiles of Random Genes found in Animals and Eukaryotes Have a Good Chance to Share Their Highly Correlated Genes of Glucose Metabolism with TP53 and PIK3CA.

As a negative control, we repeated our studies with a set of randomly selected genes from the human genome. Three independent sets of five random genes were selected from whole human genome by generating random integer numbers. Each set replaced our five representatives of Subgroups 1, 2, & 3 of GTO-Oncogenes. The phylogenetic profiles of each set of random genes and all 195 glucose metabolic genes were constructed and analyzed by the PrePhyloPro Software version 0.1.7 (Linkages Number = 100, Blast Evalve = 0.0001, Homo sapiens). For analysis of the results, we selected the correlation coefficient threshold of 0.85. The

analysis of combination of the first set of random genes with the glucose metabolic genes showed the following results [File: 7. Random Genes Set 1 & Genes of Glucose Metabolism.tar.gz]: Three of these random genes were found in prokaryotes and eukaryotes. One was found in eukaryotes, and the other one was found in animals. Among all these five genes, only one gene (PPP1R1A) had correlation of ≥ 0.85 with glucose metabolic genes. PPP1R1A was mostly found in animals, but not other eukaryotes or prokaryotes. It had correlation with 13 glucose metabolic genes: ND6, NDUFA3, NDUFC1, NDUFB4, NDUFC2-KCTD14, NDUFC2, NDUFB6, NDUFV3, NDUFA1, COX7B, COX7B2, UQCR11, & G6PC2. Nine of these genes belonged to Complex I of ETC. Two of them were from Cytochrome C Oxidase Complex (Complex IV). One was from Cytochrome bc1 Complex (Complex III), and the other one belonged to glycolysis/gluconeogenesis/pyruvate oxidation pathway.

PPP1R1A was the only gene in this set which had correlation of ≥ 0.85 with glucose metabolic genes. Since it was mostly found in animals, we compared its results with the representative of Subgroup 1, TP53. Interestingly, only one gene was found to be highly correlated with both TP53 & PPP1R1A: G6PC2. Although both genes had high correlation with subunits of Complexes I & IV, none of them were common between these two. The only shared gene belonged to glycolysis/gluconeogenesis pathway. Thus, the phylogenetic profiling results of this random gene was $\frac{1}{17}$ similar to the results of TP53.

The second analysis showed the results of the combination of Set 2 of random genes with glucose metabolic genes [File: 8. Random Genes Set 2 & Genes of Glucose Metabolism.tar.gz]. Based on these results, four out of five random genes had similarity correlation of ≥ 0.85 with glucose metabolic genes. Interestingly, three of them were highly correlated with each other too: CHRNA2, SYTL3, & OR10A2. They were found mostly in animals. CHRNA2 was highly correlated with 12 genes of glucose metabolism: COX4I2, COX7A2L, COX7A2, COX7A1, NDUFB9, NDUFB8, NDUFA7, NDUFB10, NDUFB5, ATP5J, ATP5J2, & ATP4B. In addition to these 12 genes, SYTL3 had high correlation with four other genes of glucose metabolism: COX4I1, COX6C, ATP5L, & G6PC2. Except for COX4I1, the other 15 genes were shared with OR10A2 in having high correlation. It also had high correlation with COX7C, NDUFB11, & NDUFB1. Totally, these three random genes were highly correlated with subunits of Complexes I, IV, & V. The fourth random gene, LIMS1,

was mostly found in eukaryotes. It had high correlation with five genes of glucose metabolism: ATP5F1, COX5A, COX6A1, COX6A2, & UQCRB. Thus, it seems that it was mostly correlated with Complex IV of ETC. The last random gene, ATP11A, was found in prokaryotes and eukaryotes. It did not have high correlation with any of the glucose metabolic genes.

To optimize our phylogenetic results, we decided to classify the random genes of Set 2 according to their occurrence patterns in various kingdoms and phyla. CHRNA2, SYTL3, & OR10A2 were found in animals. Thus, we compared their results with TP53, the representative of Subgroup 1. Interestingly, 15 genes of glucose metabolism were highly correlated with TP53 and ≥ 2 of the Set 2 of random genes. These genes were mostly subunits of Complexes I, IV, & V. Even ATP4B & G6PC2, the independent component of OXPHOS and an enzyme of glycolysis/gluconeogenesis pathway, respectively, were among these shared highly correlated genes. Therefore, the results of ≥ 2 of these random genes were $\frac{15}{17}$ similar to the results of TP53. The last random gene of Set 2 which had correlation of ≥ 0.85 was LIMS1. It was found in eukaryotes. Therefore, we compared it with PIK3CA, the representative of Subgroup 2. Comparison of their highly correlated genes showed that only one gene, UQCRB, was shared with these two. Thus, the results of this random gene were $\frac{1}{13}$ similar to the results of PIK3CA.

According to the third round of analysis [File: 9. Random Genes Set 3 & Genes of Glucose Metabolism.tar.gz], three of the random genes were found in eukaryotes, one was found in animals, and another one was found in prokaryotes and eukaryotes. Four of these genes had correlation of ≥ 0.85 with glucose metabolic genes. Such as the second analysis, three random genes were highly correlated with each other: TRIM68, STON1, & UBE4B. Phylogenetic profiling comparison showed that 10 genes of glucose metabolism had high correlation with TRIM68: COX6B1, COX5B, COX17, UQCRB, NDUFA6, HK1, HK2, HK3, GCK, & HKDC1. The next random gene, STON1, had also high correlation with these genes. In addition to them, it had high correlation with COX6B2, NDUFA8, NDUFA2, & NDUFA13. Except for NDUFA2, the other 13 genes of glucose metabolism had high correlation with UBE4B. In addition to subunits of Complexes I & IV, these three random genes had high correlation with Hexokinases too. The fourth random gene which had high correlation with glucose metabolic genes was DPCD. It was mostly found in animals. It

had correlation of ≥ 0.85 with five genes of glucose metabolism: ATP5J, ATP4B, COX7A2L, COX7A2, & COX7A1. Thus, it seems that this random gene had high correlation with subunits of Complex IV of ETC.

The three highly correlated random genes of Set 3, TRIM68, STON1, & UBE4B, were found in eukaryotes. Because of that, we compared their results with PIK3CA: the representative of Subgroup 2. 13 glucose metabolic genes which were highly correlated with PIK3CA were shared with ≥ 2 of these random genes. Thus, the highly correlated genes with at least two of these random genes were $\frac{13}{13}$ similar to the genes highly correlated with PIK3CA. We also compared the results of DPCD with TP53's. Interestingly, all the highly correlated genes with DPCD were shared with TP53's results. Because it had five highly correlated genes from glucose metabolism and TP53 had 17, $\frac{5}{17}$ of the results of TP53 could be predicted by this random gene.

We performed three round of phylogenetic analysis with combination of 5 random genes and 195 glucose metabolic genes. The purpose of these analyses was to find about the chances of getting the same correlated genes resulted from the representatives of Subgroups 1, 2 & 3, via selecting random genes. Totally, we analyzed 15 random genes. Five of these random genes were found in prokaryotes and eukaryotes (3 from first + 1 from second + 1 from third analyses). Interestingly, none of them had high correlation (≥ 0.85) with the aforementioned metabolic genes. It could show that the genes belonged to this category, were highly conserved. Thus, there was a very little chance for them to share their highly correlated genes with other genes (inputs). We had also five random genes found in eukaryotes (1 from first + 1 from second + 3 from third analyses). No high correlation was observed for the only random gene of this category from the first round of analysis. For the second analysis, the single gene of this category resulted in high correlation with genes that were shared with $\frac{1}{13}$ of the corresponding results of PIK3CA. The same calculations for the three random genes of this category from the third analysis resulted in $\frac{13}{13}$ of the shared glucose metabolic genes highly correlated with PIK3CA. These results could show that there was a higher chance for random genes found in eukaryotes to get high correlated genes similar to oncogenes of Subgroup 2 rather than the random genes found in both prokaryotes and eukaryotes. There were also five random genes found in animals (1 from first + 3 from second +

1 from third analyses). Interestingly, all of them had high correlation with the glucose metabolic genes. Totally, $\frac{1}{17}$, $\frac{15}{17}$, & $\frac{5}{17}$ of glucose metabolic genes which were highly correlated with TP53 were similar to the corresponding results of these random genes in the first, second, and third round of analysis, respectively. Finally, we could see that the random genes found in animals had the highest chance to be highly correlated with the genes that were shared with oncogenes of Subgroup 1.

According to the phylogenetic results obtained from analysis of random genes, we could conclude that the chance of selecting a random gene found in each category of animals, eukaryotes, and both prokaryotes and eukaryotes, were almost equal. This chance was 33.3% ($\frac{5}{15}$) for each category. Despite this result, we need to analyze all the human genes to be able to generalize this chance to other randomly selected genes. However, the chance of obtaining similar phylogenetic results from random genes was not even near equal. This chance for two independent genes (inputs) found in different categories was the lowest, particularly when one input was found in animals and the other one was found in both prokaryotes and eukaryotes. Because of that, we categorized different genes according to their occurrence pattern, and then compared their phylogenetic profiling. Thus, this chance increased when both inputs were from one category. Based on our calculations of random genes, inputs found in animals had the highest chance to produce similar results (i.e. High Chance of False Positivity). On the other hand, this chance was lowest for genes found in both prokaryotes and eukaryotes (Low Chance of False Positivity). This similarity in the results of random genes, if they were really irrelative to the target genes, could increase false positive outputs. I think the chance for this similarity maximizes in genes found in animals, because the number of animal genomes used as the reference taxa in the phylogenetic profiling was much fewer than eukaryotes and prokaryotes. To improve our results, and eliminate false positivity, we need more sequenced genomes from eukaryotes and animals. As the metabolic genes, particularly core components of metabolic pathways, are usually conserved during evolution, there is a chance that they would show correlation with a high number of genes that have a similar pattern of conservation, due to their essential roles in an entire kingdom.

Novel connections between oncogenes and glucose metabolism

Another goal of this study was to identify new connections between oncogenes and the glucose metabolic pathway that may drive cancer cell proliferation. To achieve this goal, we utilized phylogenetic profiling to predict the correlation of oncogenes, not implicated metabolism, by comparing their phylogenetic profiles with the GTO-Oncogenes known to have a role in cellular metabolism, such as TP53, PTEN, PIK3CA, BRAF & EGFR. Our investigation of the correlation of phylogenetic profiles of each of the aforementioned oncogenes with genes, that were part of the metabolic pathways, showed that TP53 & PIK3CA, but not BRAF, EGFR & PTEN, were highly correlated with some components of OXPHOS and glycolysis/gluconeogenesis pathways. After analyzing phylogenetic profiles of random genes, their results were compared with the GTO-Oncogenes. It was estimated that genes found in animals & eukaryotes had a higher chance to correlate with the other genes found in the same phylum or kingdom. This estimation did not disprove any correlation of TP53 or PIK3CA with the genes in the metabolic pathways, but we had to be aware about the possible false positive results of these subgroups of GTO-Oncogenes. We, therefore, used the five representatives of Subgroups 1, 2 & 3 as the reference oncogene sets to predict a potential role for oncogenes in cellular metabolism.

To narrow down the number of oncogenes we considered the number of occurrences of mutations (Repetition Number) for each oncogene in different cancers based on the data from (Lawrence et al., 2014) [File: 1. Supplementary Table 3 (Lawrence et al., 2014).xlsx]. According to our calculations [File: 2. List of Oncogenes.xlsx], mutations of 135 oncogenes were found only in one type of cancer. 50 oncogenes were mutated in two types of cancers. Mutations of 18 oncogenes were found in three types of cancer, and 30 oncogenes were found in ≥ 4 types of cancers. 11 out of 17 GTO-Oncogenes were the oncogenes with Repetition Number ≥ 2 . For this study, we considered the oncogenes with ≥ 2 repetitions; excluding the 11 GTO-Oncogenes, a total of 87 oncogenes were included. The phylogenetic profiles of these 87 oncogenes as well as the five representatives of Subgroups 1, 2, & 3, were constructed and examined by the PrePhyloPro Software version 0.1.7 (Linkages Number = 100, Blast Evaluated = 0.0001, Homo sapiens) [File: 10. Functional Prediction of Unknown Oncogenes.tar.gz]. At first glance on “Correlation Matrix Plot” of the results (Figure 10), it was obvious that most of the analyzed

oncogenes were clustered into 3 subgroups: 1) Subgroup 1, which were found in animals. 2) Subgroup 2, which were found in eukaryotes. 3) Subgroup 3, which were found in both prokaryotes & eukaryotes. Dissimilar to plots of glucose metabolic genes, “Correlation Matrix Plot” of these oncogenes were full of red (correlation coefficient >0.75), orange (correlation coefficient $0.50-0.75$), and white spots (correlation coefficient $\sim 0.35-0.50$). It seemed that some of the oncogenes within each subgroup were highly correlated with each other. For selecting the highly correlated oncogenes, the similarity correlation of 0.85 was adjusted as the cut off value. TP53 had similarity correlation ≥ 0.85 with 12 oncogenes with the average correlation value of 0.95: EZR, APC, MGA, ERBB3, PIK3R1, NFE2L2, ANK3, RUNX1, CASP8, ELF3, SMAD4, & RXRA. Using this threshold, 15 oncogenes were highly correlated with PTEN with the average score of 0.86: CTNNB1, RASA1, SMARCB1, ARHGAP35, STAG2, RBM10, CTCF, ZNF471, U2AF1, PBRM1, BRCA1, CREBBP, HIST1H3B, EP300, & ATM. In addition to them, PIK3CA had high correlation with 7 oncogenes with the average value of 0.93: SETDB1, NSD1, ARID2, PTPN11, NCOR1, RAC1, & SPEN. BRAF had high correlation with 9 oncogenes with the average score of 0.90: CDK12, MAP3K1, STK11, MAP2K4, MAP2K1, TGFB2, EPHA2, FGFR2, & FGFR3. At this threshold, EGFR had correlation with 8 oncogenes, which 6 of them were shared with BRAF: CDK12, MAP3K1, TGFB2, EPHA2, FGFR2, FGFR3, KIT, & FLT3. The average of correlation value for EGFR was 0.88. Near half of the initial subset, i.e. a total of 45 oncogenes, were considered as candidates involved in by this threshold (i.e. ≥ 0.85).

The use of prokaryotes has been suggested to be preferable to eukaryotes as the reference genomes for phylogenetic profiling of components of pathways involved in lipid or carbohydrate metabolism (Jothi et al., 2007). To further narrow down the list of oncogenes, therefore, we considered the 11 candidates of Subgroup 3 which were highly linked with BRAF & EGFR. As these candidates were found in both prokaryotes and eukaryotes, they could be preferably conserved during evolution, and have interactions with BRAF & EGFR.

Although we suggested 11 candidates from Subgroup 3 to be correlated with regulatory oncogenes involved in glucose metabolic shifts in cancer cells proliferation (i.e. GTO-Oncogenes), we wanted to introduce the best candidates from all the three subgroups. Such as selecting the representatives of subgroups of GTO-Oncogenes, we analyzed more data to reduce the number

of 45 oncogenes. The number of occurrence (Repetition Number) and the average of the mutation rate of all these 45 oncogenes were calculated based on the data from (Lawrence et al., 2014). The mitochondrial localization and gene position of them were also included [File: 11. Additional Analysis of Candidate Oncogenes.xlsx]. According to these results, only one of these 45 oncogenes was considered as mitochondrial protein: CASP8. By selecting 10.0% as the threshold for mutation rate, we classified them into High & Low Mutated Candidate Oncogenes. (Table 2). By considering high mutated oncogenes, 7 candidate oncogenes were selected.

To optimize our functional predictions, we needed to compare the above phylogenetic profiles with the phylogenetic profiles of random genes. The aforementioned 15 random genes we had analyzed for investigating their correlation with glucose metabolic genes were used for this step too. They were checked to be oncogene or not. None of these random genes were found in our list of oncogenes. The phylogenetic profiles of these 15 random genes as well as 45 candidate oncogenes, and 5 representatives of subgroups (total 65 gene inputs) were constructed and analyzed by the PrePhyloPro Software version 0.1.7 (Linkages Number = 100, Blast Evalue = 0.0001, Homo sapiens) [File: 12. Functional Prediction of Random Genes.tar.gz]. At threshold of 0.85, seven random genes did not have correlation with any of the representatives of subgroups 1-3, and their highly correlated oncogenes: ATP11A, ARSD, BRIP1, KRIT1, PPP1R1A, SRSF5, & CACNA1E. Four random genes had high correlation with TP53 and its correlated candidates: DPCD, CHRNA2, SYTL3, & OR10A2. DPCD had high correlation with $\frac{3}{12}$ of the high correlated candidates of TP53. The other three were highly correlated with $\frac{12}{12}$ of the predicted candidates of TP53. One random gene, LIMS1, had correlation with $\frac{7}{15}$ of high correlated candidates of PTEN. Three random genes, TRIM68, STON1, & UBE4B, were correlated with $\frac{20}{22}$, $\frac{21}{22}$, & $\frac{22}{22}$ of the predicted candidates of PIK3CA. According to these results, more than half of our random genes could predict the highly correlated candidates in Subgroups 1 & 2 (found in Animals or Eukaryotes, respectively), but not Subgroup 3 (found in both Prokaryotes & Eukaryotes). Due to low rate of high correlation in random genes of Subgroup 3, we could conclude that the predicted results of BRAF and EGFR were the best candidates to function in regulation of glucose metabolic shifts in cancer cells.

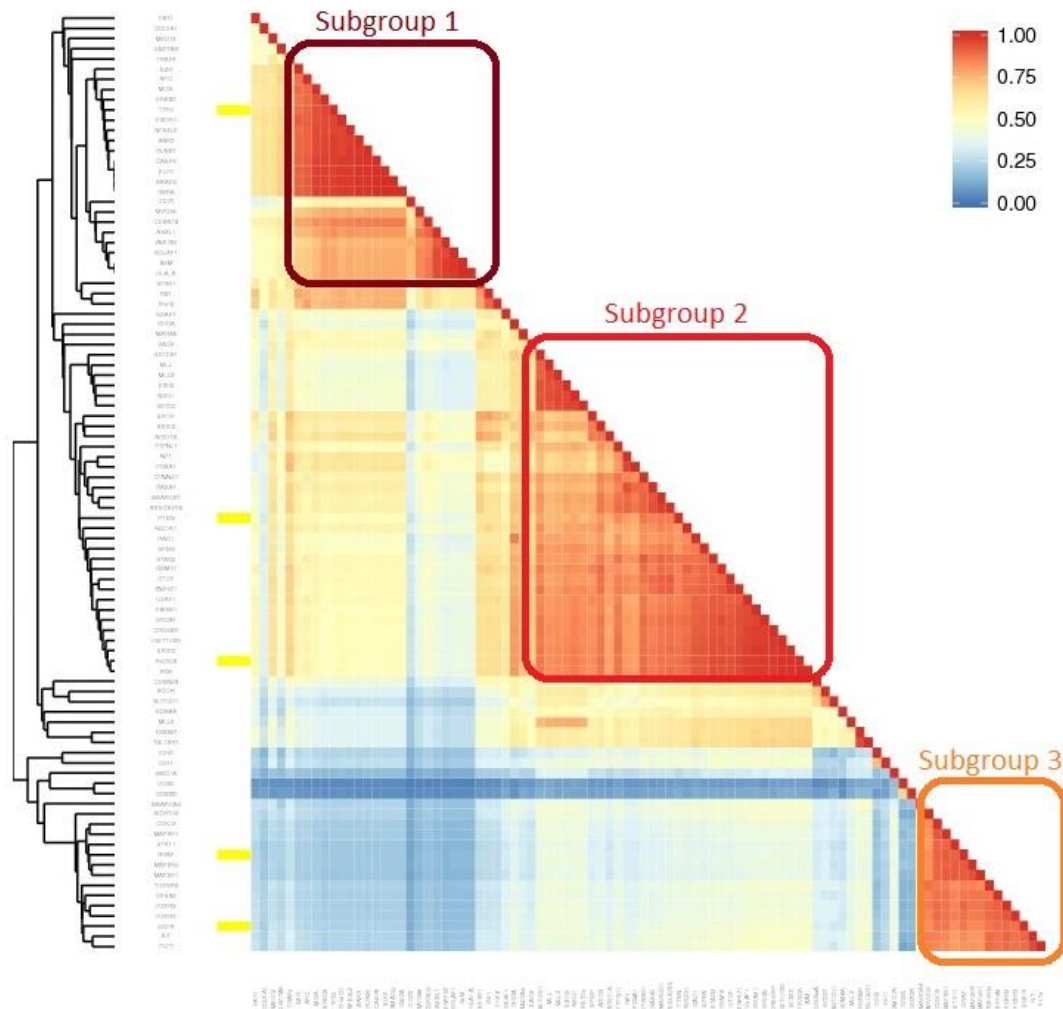


Figure 10. Correlation Matrix Plot of 87 Selected Oncogenes and Representatives of Subgroups 1, 2 & 3 of GTO-Oncogenes.

It is clear that phylogenetic profiles of most of the selected oncogenes preferred to cluster together. Similar to GTO-Oncogenes such as TP53, PTEN, etc, the selected oncogenes were categorized into 3 subgroups, based on their occurrence patterns in different kingdoms & phyla. Except for a few oncogenes, most of the oncogenes of each subgroup were highly correlated with each other.

	Subgroup 1 (found in Animals)	Subgroup 2 (found in Eukaryotes)	Subgroup 3 (found in Prokaryotes & Eukaryotes)
High Mutation rate ($\geq 10.0\%$)	APC, PIK3R1, ANK3	CTNNB1, PBRM1, CREBBP	FLT3
Low Mutation rate ($< 10.0\%$)	EZR, MGA, ERBB3, NFE2L2, RUNX1, CASP8, ELF3, SMAD4, RXRA	RASA1, SMARCB1, ARHGAP35, STAG2, RBM10, CTCF, ZNF471, U2AF1, BRCA1, HIST1H3B, EP300, ATM, SETDB1, NSD1, ARID2, PTPN11, NCOR1, RAC1, SPEN	CDK12, MAP3K1, STK11, MAP2K4, MAP2K1, TGFBR2, EPHA2, FGFR2, FGFR3, KIT

Table 2. Classification of 45 Candidate Oncogenes Based on Their Mutation Rate.

Similar to GTO-Oncogenes, candidate oncogenes with mutation rate of $\geq 10.0\%$ or $< 10.0\%$ were classified as High or Low Mutated Oncogenes, respectively.

Discussion & Conclusion:

A hallmark of cancer is the metabolic reprogramming of glucose consumption and bioenergetics. This effect seems to be regulated by the oncogenes we have called the GTO-Oncogenes. In this study, we searched for these oncogenes, investigated their phylogenetic correlations with all the genes of glycolysis, TCA cycle, and OXPHOS, and found other oncogenes contributing to this metabolic reprogramming via phylogenetic profiling comparison. In this section, we discussed the limitations of phylogenetic profiling, especially false positive results, and the experimental data to support our predictive phylogenetic results.

The phylogenetic profiling comparison of GTO-Oncogenes demonstrated high correlation among PIK3CA, PTEN, and mTOR, but not Akt, while all of them belonged to a signaling pathway (DeBerardinis et al., 2008). Also, although Akt could be activated by Ras family of oncogenes (Elstrom et al., 2004), no significant correlation was observed between Akt and the oncogenes of the Ras family. Instead, other signaling molecules, such as EGFR and HER2/neu, which could activate some components of PI3K/Akt/mTOR pathway (DeBerardinis et al., 2008), had high

correlation only with Akt. Perhaps this is due to the fact that Akt could be activated in different signaling pathways. Another scenario suggests separate evolutionary routes for the components of metabolic pathways. Each metabolic pathway could have two components: a highly conserved set of core genes which makes up a small fraction of the pathway, and a large subset of newly evolved genes added to the core pathway. These newly evolved genes result in adaptation of organism to new conditions (Muley and Ranjan, 2013). The highly conserved and newly evolved portions both could form different evolutionary modules. The genes in each module have higher co-evolutionary linkages with each other. Not all the cellular pathways and networks have evolutionary modules. It is estimated that about half of the protein complexes or metabolic networks have meaningful evolutionary modules. To unravel more asymmetric linkages within different modules of a pathway, more sequenced genomes are required to improve the resolution of phylogenetic profiles (Kensche et al., 2008).

Due to the differences in the phylogenetic profiles of the members of the PI3K/Akt/mTOR pathway, one approach would be to exclude all the oncogenes in each subgroup that only had signaling or transcriptional activity and focus on metabolic proteins. We decided against this approach, since except for Subgroup 3, all the members of the other subgroups did not show very high correlation within their clusters. Instead, we selected oncogenes with high mutation rate. Among highly mutated GTO-Oncogenes, we selected the oncogenes with high correlation in their subgroups: TP53 from Subgroup 1, PTEN and PIK3CA from Subgroup 2, and EGFR and BRAF from Subgroup 3. Although none of these oncogenes are mitochondrial, some of the protein variants, such as wild-type TP53 (Vaseva and Moll, 2013), PTEN (Zhu et al., 2006), some activating mutants of BRAF (Lee et al., 2011), and EGFR (Demory et al., 2009), can bind or enter the mitochondria under certain conditions. Because of their potential presence in mitochondria, we could expect to see high correlation of their phylogenetic profiles with some components of the TCA cycle and OXPHOS. Therefore, to identify potential interactions of glucose metabolic genes with oncogenes, phylogenetic profiles of all genes involved in glycolysis, TCA cycle, and OXPHOS were constructed and compared with these selected GTO-Oncogenes.

According to primary phylogenetic profile analysis, TP53 showed a high correlation with some subunits of mitochondrial complexes 1, 4, and 5. It also had high correlation with one

enzyme of glycolysis/gluconeogenesis pathway and an independent component of OXPHOS. Unfortunately, no experimental evidences of their direct interaction with TP53 were found in the literature, except for the role of TP53 in synthesis of Cytochrome C Oxidase subunit 2 (SCO2 or COX2). As it was mentioned before, SCO2 is an important subunit for Complex IV function, and its expression is regulated by TP53 (Wanka et al., 2012). In addition, an indirect interaction of TP53 with ATP synthase was found. Strap (stress-responsive activator of p300), the transcriptional coactivator of TP53, can also be transferred into mitochondria, can bind to ATP synthase and inhibit its activity (Maniam et al., 2015).

Phylogenetic profile of PIK3CA had high correlation with mitochondrial complexes 1, 3, and 4, as well as Hexokinase family of glycolytic enzymes. Phylogenetic profile of PTEN was not significantly correlated with aforementioned glucose metabolic genes. Unfortunately, there was no evidence of direct interaction of PTEN and PIK3CA with these genes of glucose metabolism.

Interestingly based on the primary results of phylogenetic profiling, no significant correlation was observed between EGFR and BRAF with all the 195 genes of glucose metabolism. These results were against our expectation to observe high correlation of these oncogenes with highly conserved pathways of glucose metabolism. Glucose metabolic genes found in both prokaryotes and eukaryotes preferred to form phylogenetic clusters with each other in small numbers (Figure 9). These clusters could be a clue for detecting phylogenetically conserved pathways. Unfortunately, neither BRAF nor EGFR had a significant correlation with these clusters.

To optimize the primary results of phylogenetic profiling, it was required to compare them with the results of negative controls. There was little chance for randomly selected genes to be functionally correlated. They were suggested to be used as negative controls (Pellegrini et al., 1999). However, for negative controls, we could also use oncogenes that certainly were not involved in glucose metabolism of cancer cells. Due to complexity of biological systems and uncharacterized pathways, finding oncogenes unrelated to the metabolic pathways seemed very difficult. Thus, the results of randomly selected genes were prepared and analyzed to optimize the primary main results. Among all 15 randomly selected genes, each subgroup of animals, eukaryotes, and both prokaryotes and eukaryotes had 5 genes. Similar to the classification of

oncogenes, it might be helpful to categorize all the human genes according to their occurrence pattern. Then, we could calculate the probability of selecting random genes from each subgroup. Perhaps it would be needed to select more random genes from phylum of animals or kingdom of eukaryotes. Nevertheless, analysis of randomly selected genes found in prokaryotes & eukaryotes showed no significant correlation with EGFR and BRAF, or their highly correlated metabolic genes. It could demonstrate that genes of this subgroup belong to highly conserved pathways or complexes. Therefore, there was a low probability for oncogenes of Subgroup 3 to share their highly correlated genes (outputs) with randomly selected genes of this subgroup. The phylogenetic results of random genes found in eukaryotes or animals showed that the probability of predicting highly correlated genes of oncogenes of subgroups 1 and 2 with random genes are higher than the previous subgroup. In other words, the order of false positivity of random genes is: Animals > Eukaryotes > Both Prokaryotes and Eukaryotes. One reason for this order could be the value of conservation of pathways or complexes. From the evolutionary perspective, genes found only in animals are considered younger than the genes found in eukaryotes or both prokaryotes and eukaryotes. Therefore, they are more related to each other, and their predictions could have more similarity. The other reason which might be more important is the number of sequenced genomes. The phylogenetic profiles were prepared from 276 eukaryotic and 696 fully sequenced prokaryotic genomes. It seemed that ~33% of the eukaryotic genomes were from animals (91 sequenced genomes of animals). Since the number of reference animal species was much fewer than the eukaryotes or both prokaryotes and eukaryotes, the discrimination of phylogenetic profiles of genes found in animals became harder as they showed more similarity. Subsequently the probability of acquiring false positive results was higher for this subgroup. To enhance the accuracy of predictive results of this group and reduce false positivity, more sequenced genomes from eukaryotes and especially animals are required. The other approach to reduce false positive results, is "Boostrapping". This statistical method is very useful to calculate false positivity of random samples (Efron, 1979). For this purpose, bootstrapping needs to be incorporated into the PrePhyloPro Software version in subsequent versions.

Due to the importance of glucose metabolism in development of cancer cell, metabolic pathways involved in metabolic reprogramming of cancer cells are considered as a new target

for cancer treatment. They also could be helpful in understanding the correlation of other metabolic diseases, such as diabetes, with cancer. Discovery of the underlying biological networks involved in the metabolic diseases and cancer can support repurposing of drugs for cancer treatment (Pierotti et al., 2013). In the last part of the results, we tried to predict function of other oncogenes that could be correlated with reprogramming of glucose metabolism in cancer cell proliferation via constructing and comparing their phylogenetic profiles with BRAF, EGFR, PTEN, PIK3CA, and TP53. A total number of 45 oncogenes with Repetition Number ≥ 2 were highly correlated with these 5 GTO-Oncogenes. Analysis of the phylogenetic profiles of these 5 oncogenes with genes of glucose metabolism have shown that the false positive results of oncogenes of subgroups 1 and 2, such as TP53, PTEN and PIK3CA, could be higher than the oncogenes of Subgroup 3. In addition, prokaryotic genomes were suggested as the best reference taxa for phylogenetic profiling of components of carbohydrate or lipid metabolism (Jothi et al., 2007). Thus, we included the 11 candidate oncogenes that had high correlation with BRAF and EGFR: CDK12, MAP3K1, STK11, MAP2K4, MAP2K1, TGFBR2, EPHA2, FGFR2, FGFR3, KIT, and FLT3. Based on the phylogenetic profiling results, we suggest that these oncogenes are involved in regulation of glucose metabolic shift in cancer cells through the interactions with the GTO-Oncogenes of Subgroup 3; oncogenes that were found in both prokaryotes and eukaryotes.

To confirm these studies, we looked for evidence in the literature, preferably experimental studies, for connections of these 11 candidate oncogenes with glucose metabolic pathways. We found that a few of these candidates, such as MAP3K1, STK11, and MAP2K1, were shown to be fully or partially involved in changes of cellular glucose metabolism, e.g. OXPHOS. Activation of MAP3K1 by sodium nitroprusside (SNP), a nitric oxide donor, results in transfer of Bax to mitochondria, reduction of activity of mitochondrial Complex 1, and decrease of ATP levels in human chondrocytes. Along with other serial events, SNP can have apoptotic effect on these cells. This apoptotic effect can be improved by overexpression of a dominant negative mutant of MAP3K1 (Cherng et al., 2008). Therefore, avoiding apoptosis is another hallmark of cancer promoted by MAP3K1. The other candidate, STK11, is a tumor suppressor gene encoding the liver serine/threonine kinase B1 (LKB1). It is critically involved in regulation of glucose metabolism. In breast cancer cells, loss of LKB1 results in metabolic reprogramming of cancer cells in an

HER2/neu-dependent manner by activation of Akt/mTOR signaling (Dupuy et al., 2013). Inactivation of STK11 in zebrafish increases the level of pyruvate dehydrogenase kinase 2 (PDK2). Subsequently, PDK2 phosphorylates and inactivates pyruvate dehydrogenase complex, leading to stimulation of glycolysis as well as lactate production. It is suggested that the mutants of STK11 could be applied to identify drugs targeting metabolic reprogramming of cancer cells (Kuang et al., 2016). Also, loss of STK11 in mouse embryonic fibroblasts enhances uptake and consumption of glucose and glutamine via deregulation of cell bioenergetics and metabolism in a HIF1 α -dependent manner. In tumor cells, loss of STK11 stimulates glycolysis and TCA cycle activity as well as increases the expression of HIF1 α in presence of normal concentration of oxygen. It is also critically involved in control of cell metabolism and growth by regulating mTOR (Faubert et al., 2014). In human cancer cells expressing BRAF^{V600E} mutant, MAP2K1 (MEK1) could be activated by BRAF^{V600E} mutant to trigger MEK-ERK signaling pathway (Kang et al., 2015). As mentioned before, the Ras-BRAF-MEK-ERK pathway contributes to regulation of many hallmarks of cancer, such as cell growth, survival, invasion, angiogenesis, and the metabolic reprogramming of the cells (Kumar et al., 2007, Hall et al., 2013). BRAF^{V600E} mutant could also enhance expression of 3-hydroxy-3-methylglutaryl-CoA lyase (HMGCL) via an octamer transcription factor Oct-1. HMGCL is involved in ketogenesis pathway, and its silencing results in inhibition of growth and proliferation of human melanoma cells. Interestingly, acetoacetate, the product of this enzyme, stimulates direct interaction of BRAF^{V600E} mutant with MAP2K1. Thus, this constitutive activating mutant couples MEK-ERK signaling to metabolic ketogenesis pathway (Kang et al., 2015).

No experimental studies directly talked about EPHA2 or KIT implications in reprogramming of glucose metabolism in cancer cells, but a study showed that EPHA2 stimulation by ephrin-A1 results in inhibition of Akt activity in human corneal epithelial cells (Kaplan et al., 2012). In another study, it was found that glucose metabolism in human corneal epithelial keratinocytes could be changed through the activation of Akt/Glycogen Synthase Kinase (GSK)-3 β pathway by KIT signaling (Peng et al., 2013). Among all 11 candidate oncogenes, at least five of them (MAP3K1, STK11, MAP2K1, EPHA2 and KIT) were found to be involved in regulation of metabolic pathways of cancer cells. Interestingly, most of these oncogenes trigger pathways that are common in one important effector molecule, Akt. According to phylogenetic

profiling results, Akt1, 2, and 3, belong to Subgroup 3. Thus, it seems that Akt is a great target for modulating the metabolic pathways connected to the proliferation of cancer cells. The remaining 6 candidate oncogenes may also play a role in this process. Experimental data are required to elucidate the mechanism of their involvement.

To overcome the high false positive rate of the representatives of subgroups 1 and 2, and to narrow down the number of 34 candidate oncogenes, we calculated the average mutation rates of these genes. Highly mutated candidate oncogenes were: APC, PIK3R1, and ANK3 for Subgroup1. CTNNB1, PBRM1, and CREBBP for Subgroup 2. Previous evidence showed that APC and CTNNB1 could participate in deregulation of glucose metabolism in cancer cells. Interestingly, both these genes belong to the Wnt/beta-catenin signaling pathway. This signaling pathway is involved in reprogramming of glucose metabolism in colon cancer cells (Pate et al., 2014). CTNNB1 encodes beta-catenin, a critical component of this signaling cascade. Adenomatous polyposis coli (APC) is a tumor suppressor, involved in regulation of Wnt/beta-catenin signaling pathway. It binds to beta-catenin to inhibit the signaling pathway. Downregulation of APC by glucose regulated protein 78 (GRP78) results in release of beta-catenin and activation of this signaling pathway leading to a metabolic shift in glucose metabolism (Li et al., 2014). Acute activation of beta-catenin occurs after deletion of APC in liver cells of mice. It results in downregulation of ATP5a1 and ATP5b, the two catalytic subunits of mitochondrial Complex 5, as well as stimulation of lactate dehydrogenase (LDH) (Chafey et al., 2009).

Optimization of the phylogenetic results of the GTO-Oncogenes with randomly selected genes revealed the likelihood of high false positivity rate for the oncogenes of subgroups 1 and 2. Since the genetic pool and tissue of origin are two important factors determining the metabolic features of cancer (Yuneva et al., 2012), future studies should focus on a particular type and original tissue of cancer to exclude some of false positive results.

In this study, we looked for phylogenetically linked genes of glucose metabolism with some of the most important GTO-Oncogenes, such as TP53, PTEN, PIK3CA, BRAF and EGFR. We realized that the GTO-Oncogenes found in both prokaryotes and eukaryotes seem to belong to highly conserved pathways which do not have significant phylogenetic linkages with glucose

metabolism genes. On the other hand, GTO-Oncogenes found in eukaryotes or animals prefer to have high correlation with some subunits of Complexes of OXPHOS rather than components of TCA cycle or glycolysis. Interestingly, we observed that the rate of false positivity of the GTO-Oncogenes depends on their occurrence patterns in different kingdoms and phyla. Oncogenes found in animals have the highest, while oncogenes found in both prokaryotes and eukaryotes have the lowest rate of false positivity.

To conclude, we predicted functions of other oncogenes in metabolic reprogramming of cancer cells by phylogenetic profiling comparison with TP53, PTEN, PIK3CA, BRAF, and EGFR. Due to low rate of false positivity and highly conserved pathways, we suggested that the oncogenes with phylogenetic linkages with BRAF and EGFR could be critically interconnected in the signaling and metabolic pathways participating in glucose metabolic shift in cancer cells. Their shared effectors, such as Akt, therefore, could be potent targets for modulation of cancer metabolism and for treatment of cancer.

References:

- Adams JM, Cory S (2007) The Bcl-2 apoptotic switch in cancer development and therapy. *Oncogene* 26:1324-1337.
- Arismendi-Morillo G (2009) Electron microscopy morphology of the mitochondrial network in human cancer. *Int J Biochem Cell Biol* 41:2062-2068.
- Basu MK, Selengut JD, Haft DH (2011) ProPhylo: partial phylogenetic profiling to guide protein family construction and assignment of biological process. *BMC Bioinformatics* 12:434.
- Bensaad K, Tsuruta A, Selak MA, Vidal MN, Nakano K, Bartrons R, Gottlieb E, Vousden KH (2006) TIGAR, a p53-inducible regulator of glycolysis and apoptosis. *Cell* 126:107-120.
- Bost F, Decoux-Pouillot AG, Tanti JF, Clavel S (2016) Energy disruptors: rising stars in anticancer therapy? *Oncogenesis* 5:e188.
- Calvo SE, Clauser KR, Mootha VK (2016) MitoCarta2.0: an updated inventory of mammalian mitochondrial proteins. *Nucleic Acids Res* 44:D1251-1257.
- Chafey P, Finzi L, Boisgard R, Cauzac M, Clary G, Broussard C, Pegorier JP, Guillonnet F, Mayeux P, Camoin L, Tavitian B, Colnot S, Perret C (2009) Proteomic analysis of beta-catenin activation in mouse liver by DIGE analysis identifies glucose metabolism as a new target of the Wnt pathway. *Proteomics* 9:3889-3900.
- Cheng Y, Perocchi F (2015) Prediction of mitochondrial protein function by comparative physiology and phylogenetic profiling. *Methods Mol Biol* 1264:321-329.
- Cherng YG, Chang HC, Lin YL, Kuo ML, Chiu WT, Chen RM (2008) Apoptotic insults to human chondrocytes induced by sodium nitroprusside are involved in sequential events, including cytoskeletal remodeling, phosphorylation of mitogen-activated protein kinase kinase kinase-1/c-Jun N-terminal kinase, and Bax-mitochondria-mediated caspase activation. *J Orthop Res* 26:1018-1026.
- Choi SS, Cha SH, Tappert C (2010) A Survey of Binary Similarity and Distance Measures. *Journal of Systemics, Cybernetics and informatics* 8:43-48.
- Cui J, DeLuca TF, Jung JY, Wall DP (2011) Phylogenetically informed logic relationships improve detection of biological network organization. *BMC Bioinformatics* 12:476.
- Darlington CD (1948) The plasmagene theory of the origin of cancer. *Br J Cancer* 2:118-126.

- DeBerardinis RJ, Lum JJ, Hatzivassiliou G, Thompson CB (2008) The biology of cancer: metabolic reprogramming fuels cell growth and proliferation. *Cell Metab* 7:11-20.
- Demory ML, Boerner JL, Davidson R, Faust W, Miyake T, Lee I, Huttemann M, Douglas R, Haddad G, Parsons SJ (2009) Epidermal growth factor receptor translocation to the mitochondria: regulation and effect. *J Biol Chem* 284:36592-36604.
- Dey G, Jaimovich A, Collins SR, Seki A, Meyer T (2015) Systematic Discovery of Human Gene Function and Principles of Modular Organization through Phylogenetic Profiling. *Cell Rep.*
- Dey G, Meyer T (2015) Phylogenetic Profiling for Probing the Modular Architecture of the Human Genome. *Cell Syst* 1:106-115.
- Dudek H, Datta SR, Franke TF, Birnbaum MJ, Yao R, Cooper GM, Segal RA, Kaplan DR, Greenberg ME (1997) Regulation of neuronal survival by the serine-threonine protein kinase Akt. *Science* 275:661-665.
- Dupuy F, Griss T, Blagih J, Bridon G, Avizonis D, Ling C, Dong Z, Siwak DR, Annis MG, Mills GB, Muller WJ, Siegel PM, Jones RG (2013) LKB1 is a central regulator of tumor initiation and pro-growth metabolism in ErbB2-mediated breast cancer. *Cancer Metab* 1:18.
- Efron B (1979) Bootstrap Methods: Another Look at the Jackknife. 1-26.
- Elliott RL, Jiang XP, Head JF (2012) Mitochondria organelle transplantation: introduction of normal epithelial mitochondria into human cancer cells inhibits proliferation and increases drug sensitivity. *Breast Cancer Res Treat* 136:347-354.
- Elstrom RL, Bauer DE, Buzzai M, Karnauskas R, Harris MH, Plas DR, Zhuang H, Cinalli RM, Alavi A, Rudin CM, Thompson CB (2004) Akt stimulates aerobic glycolysis in cancer cells. *Cancer Res* 64:3892-3899.
- Fabregat I, Vitorica J, Satrustegui J, Machado A (1985) The pentose phosphate cycle is regulated by NADPH/NADP ratio in rat liver. *Arch Biochem Biophys* 236:110-118.
- Faubert B, Vincent EE, Griss T, Samborska B, Izreig S, Svensson RU, Mamer OA, Avizonis D, Shackelford DB, Shaw RJ, Jones RG (2014) Loss of the tumor suppressor LKB1 promotes metabolic reprogramming of cancer cells via HIF-1 α . *Proc Natl Acad Sci U S A* 111:2554-2559.
- Fiske BP, Vander Heiden MG (2012) Seeing the Warburg effect in the developing retina. *Nat Cell Biol* 14:790-791.

- Hall A, Meyle KD, Lange MK, Klima M, Sanderhoff M, Dahl C, Abildgaard C, Thorup K, Moghimi SM, Jensen PB, Bartek J, Guldberg P, Christensen C (2013) Dysfunctional oxidative phosphorylation makes malignant melanoma cells addicted to glycolysis driven by the (V600E)BRAF oncogene. *Oncotarget* 4:584-599.
- Hanahan D, Weinberg RA (2000) The hallmarks of cancer. *Cell* 100:57-70.
- Hanahan D, Weinberg RA (2011) Hallmarks of cancer: the next generation. *Cell* 144:646-674.
- Hu Y, Lu W, Chen G, Wang P, Chen Z, Zhou Y, Ogasawara M, Trachootham D, Feng L, Pelicano H, Chiao PJ, Keating MJ, Garcia-Manero G, Huang P (2012) K-ras(G12V) transformation leads to mitochondrial dysfunction and a metabolic switch from oxidative phosphorylation to glycolysis. *Cell Res* 22:399-412.
- Israel BA, Schaeffer WI (1987) Cytoplasmic suppression of malignancy. *In Vitro Cell Dev Biol* 23:627-632.
- Israel BA, Schaeffer WI (1988) Cytoplasmic mediation of malignancy. *In Vitro Cell Dev Biol* 24:487-490.
- Jiang BH, Liu LZ (2009) PI3K/PTEN signaling in angiogenesis and tumorigenesis. *Adv Cancer Res* 102:19-65.
- Jones RG, Thompson CB (2009) Tumor suppressors and cell metabolism: a recipe for cancer growth. *Genes Dev* 23:537-548.
- Jothi R, Przytycka TM, Aravind L (2007) Discovering functional linkages and uncharacterized cellular pathways using phylogenetic profile comparisons: a comprehensive assessment. *BMC Bioinformatics* 8:173.
- Kaipappattu BA, Ma Y, Park JH, Lee TL, Zhang Y, Yotnda P, Creighton CJ, Chan WY, Wong LJ (2013) Crosstalk from non-cancerous mitochondria can inhibit tumor properties of metastatic cells by suppressing oncogenic pathways. *PLoS One* 8:e61747.
- Kanehisa M, Sato Y, Kawashima M, Furumichi M, Tanabe M (2016) KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Res* 44:D457-462.
- Kang HB, Fan J, Lin R, Elf S, Ji Q, Zhao L, Jin L, Seo JH, Shan C, Arbiser JL, Cohen C, Brat D, Miziorko HM, Kim E, Abdel-Wahab O, Merghoub T, Frohling S, Scholl C, Tamayo P, Barbie DA, Zhou L, Pollack BP, Fisher K, Kudchadkar RR, Lawson DH, Sica G, Rossi M, Lonial S, Khoury HJ,

- Khuri FR, Lee BH, Boggon TJ, He C, Kang S, Chen J (2015) Metabolic Rewiring by Oncogenic BRAF V600E Links Ketogenesis Pathway to BRAF-MEK1 Signaling. *Mol Cell* 59:345-358.
- Kaplan N, Fatima A, Peng H, Bryar PJ, Lavker RM, Getsios S (2012) EphA2/Ephrin-A1 signaling complexes restrict corneal epithelial cell migration. *Invest Ophthalmol Vis Sci* 53:936-945.
- Kensche PR, van Noort V, Dutilh BE, Huynen MA (2008) Practical and theoretical advances in predicting the function of a protein by its phylogenetic distribution. *J R Soc Interface* 5:151-170.
- Kotaru AR, Shameer K, Sundaramurthy P, Joshi RC (2013) An improved hypergeometric probability method for identification of functionally linked proteins using phylogenetic profiles. *Bioinformation* 9:368-374.
- Kuang X, Liu C, Fang J, Ma W, Zhang J, Cui S (2016) The tumor suppressor gene *lkb1* is essential for glucose homeostasis during zebrafish early development. *FEBS Lett* 590:2076-2085.
- Kumar SM, Yu H, Edwards R, Chen L, Kazianis S, Brafford P, Acs G, Herlyn M, Xu X (2007) Mutant V600E BRAF increases hypoxia inducible factor-1 α expression in melanoma. *Cancer Res* 67:3177-3184.
- Lawrence MS, Stojanov P, Mermel CH, Robinson JT, Garraway LA, Golub TR, Meyerson M, Gabriel SB, Lander ES, Getz G (2014) Discovery and saturation analysis of cancer genes across 21 tumour types. *Nature* 505:495-501.
- Lee MH, Lee SE, Kim DW, Ryu MJ, Kim SJ, Kim SJ, Kim YK, Park JH, Kweon GR, Kim JM, Lee JU, De Falco V, Jo YS, Shong M (2011) Mitochondrial localization and regulation of BRAFV600E in thyroid cancer: a clinically used RAF inhibitor is unable to block the mitochondrial activities of BRAFV600E. *J Clin Endocrinol Metab* 96:E19-30.
- Li L, Connelly MC, Wetmore C, Curran T, Morgan JI (2003) Mouse embryos cloned from brain tumors. *Cancer Res* 63:2733-2736.
- Li Z, Wang Y, Wu H, Zhang L, Yang P, Li Z (2014) GRP78 enhances the glutamine metabolism to support cell survival from glucose deficiency by modulating the beta-catenin signaling. *Oncotarget* 5:5369-5380.
- Lin TW, Wu JW, Chang DT (2013) Combining phylogenetic profiling-based and machine learning-based techniques to predict functional related proteins. *PLoS One* 8:e75940.

- Lord-Fontaine S, Averill-Bates DA (2002) Heat shock inactivates cellular antioxidant defenses against hydrogen peroxide: protection by glucose. *Free Radic Biol Med* 32:752-765.
- Maniam S, Coutts AS, Stratford MR, McGouran J, Kessler B, La Thangue NB (2015) Cofactor Strap regulates oxidative phosphorylation and mitochondrial p53 activity through ATP synthase. *Cell Death Differ* 22:156-163.
- Matoba S, Kang JG, Patino WD, Wragg A, Boehm M, Gavrilova O, Hurley PJ, Bunz F, Hwang PM (2006) p53 regulates mitochondrial respiration. *Science* 312:1650-1653.
- Moreira JD, Hamraz M, Abolhassani M, Bigan E, Peres S, Pauleve L, Nogueira ML, Steyaert JM, Schwartz L (2016) The Redox Status of Cancer Cells Supports Mechanisms behind the Warburg Effect. *Metabolites* 6.
- Muley VY, Ranjan A (2013) Evaluation of physical and functional protein-protein interaction prediction methods for detecting biological pathways. *PLoS One* 8:e54325.
- Neuzil J, Rohlena J, Dong LF (2012) K-Ras and mitochondria: dangerous liaisons. *Cell Res* 22:285-287.
- Osthus RC, Shim H, Kim S, Li Q, Reddy R, Mukherjee M, Xu Y, Wonsey D, Lee LA, Dang CV (2000) Deregulation of glucose transporter 1 and glycolytic gene expression by c-Myc. *J Biol Chem* 275:21797-21800.
- Pate KT, Stringari C, Sprowl-Tanio S, Wang K, TeSlaa T, Hoverter NP, McQuade MM, Garner C, Digman MA, Teitell MA, Edwards RA, Gratton E, Waterman ML (2014) Wnt signaling directs a metabolic program of glycolysis and angiogenesis in colon cancer. *EMBO J* 33:1454-1473.
- Pellegrini M (2012) Using phylogenetic profiles to predict functional relationships. *Methods Mol Biol* 804:167-177.
- Pellegrini M, Marcotte EM, Thompson MJ, Eisenberg D, Yeates TO (1999) Assigning protein functions by comparative genome analysis: protein phylogenetic profiles. *Proc Natl Acad Sci U S A* 96:4285-4288.
- Peng H, Katsnelson J, Yang W, Brown MA, Lavker RM (2013) FIH-1/c-kit signaling: a novel contributor to corneal epithelial glycogen metabolism. *Invest Ophthalmol Vis Sci* 54:2781-2786.

- Petros JA, Baumann AK, Ruiz-Pesini E, Amin MB, Sun CQ, Hall J, Lim S, Issa MM, Flanders WD, Hosseini SH, Marshall FF, Wallace DC (2005) mtDNA mutations increase tumorigenicity in prostate cancer. *Proc Natl Acad Sci U S A* 102:719-724.
- Pierotti MA, Berrino F, Gariboldi M, Melani C, Mogavero A, Negri T, Pasanisi P, Pilotti S (2013) Targeting metabolism for cancer treatment and prevention: metformin, an old drug with multi-faceted effects. *Oncogene* 32:1475-1487.
- Seyfried TN, Flores RE, Poff AM, D'Agostino DP (2014) Cancer as a metabolic disease: implications for novel therapeutics. *Carcinogenesis* 35:515-527.
- Shay JW, Werbin H (1988) Cytoplasmic suppression of tumorigenicity in reconstructed mouse cells. *Cancer Res* 48:830-833.
- Shim H, Chun YS, Lewis BC, Dang CV (1998) A unique glucose-dependent apoptotic pathway induced by c-Myc. *Proc Natl Acad Sci U S A* 95:1511-1516.
- Shim H, Dolde C, Lewis BC, Wu CS, Dang G, Jungmann RA, Dalla-Favera R, Dang CV (1997) c-Myc transactivation of LDH-A: implications for tumor metabolism and growth. *Proc Natl Acad Sci U S A* 94:6658-6663.
- Simonsen M, Maetschke SR, Ragan MA (2012) Automatic selection of reference taxa for protein-protein interaction prediction with phylogenetic profiling. *Bioinformatics* 28:851-857.
- Vaseva AV, Moll UM (2013) Identification of p53 in mitochondria. *Methods Mol Biol* 962:75-84.
- Wanka C, Brucker DP, Bahr O, Ronellenfitsch M, Weller M, Steinbach JP, Rieger J (2012) Synthesis of cytochrome C oxidase 2: a p53-dependent metabolic regulator that promotes respiratory function and protects glioma and colon cancer cells from hypoxia-induced cell death. *Oncogene* 31:3764-3776.
- Warburg O (1956) On the origin of cancer cells. *Science* 123:309-314.
- Warburg O, Wind F, Negelein E (1927) The Metabolism of Tumors in the Body. *J Gen Physiol* 8:519-530.
- Weinberg F, Hamanaka R, Wheaton WW, Weinberg S, Joseph J, Lopez M, Kalyanaraman B, Mutlu GM, Budinger GR, Chandel NS (2010) Mitochondrial metabolism and ROS generation are essential for Kras-mediated tumorigenicity. *Proc Natl Acad Sci U S A* 107:8788-8793.
- Willis SN, Adams JM (2005) Life in the balance: how BH3-only proteins induce apoptosis. *Curr Opin Cell Biol* 17:617-625.

- Yang D, Wang MT, Tang Y, Chen Y, Jiang H, Jones TT, Rao K, Brewer GJ, Singh KK, Nie D (2010) Impairment of mitochondrial respiration in mouse fibroblasts by oncogenic H-RAS(Q61L). *Cancer Biol Ther* 9:122-133.
- Yeh E, Cunningham M, Arnold H, Chasse D, Monteith T, Ivaldi G, Hahn WC, Stukenberg PT, Shenolikar S, Uchida T, Counter CM, Nevins JR, Means AR, Sears R (2004) A signalling pathway controlling c-Myc degradation that impacts oncogenic transformation of human cells. *Nat Cell Biol* 6:308-318.
- Yuneva MO, Fan TW, Allen TD, Higashi RM, Ferraris DV, Tsukamoto T, Mates JM, Alonso FJ, Wang C, Seo Y, Chen X, Bishop JM (2012) The metabolic profile of tumors depends on both the responsible genetic lesion and tissue type. *Cell Metab* 15:157-170.
- Zhu Y, Hoell P, Ahlemeyer B, Krieglstein J (2006) PTEN: a crucial mediator of mitochondria-dependent apoptosis. *Apoptosis* 11:197-207.