COMPUTATIONAL ANALYSIS OF METABOLIC REPROGRAMMING IN TUMORS

Dissertation submitted to the
Combined Faculties for the Natural Sciences and for Mathematics
of the Ruperto-Carola University of Heidelberg, Germany
for the degree of Doctor of Natural Sciences

ASHWINI KUMAR SHARMA
Division Of Theoretical Bioinformatics
Network Modelling Group
German Cancer Research Center (DKFZ)

Dissertation submitted to the
Combined Faculties for the Natural Sciences and for Mathematics
of the Ruperto-Carola University of Heidelberg, Germany
for the degree of Doctor of Natural Sciences

Presented by

Ashwini Kumar Sharma, M.Sc in Genomics
Born in: Kurseong, India
Oral Examination:

COMPUTATIONAL ANALYSIS OF METABOLIC REPROGRAMMING IN TUMORS

Referees

PROF. DR. RAINER KÖNIG

PROF. DR. PETER LICHTER

DECLARATION

I hereby declare that the submitted dissertation is my own work and has been written by myself. I have used no other sources or materials than those explicitly indicated. I have not yet presented this thesis or parts thereof to a university as part of an examination or degree. This work was carried out in the Division of Theoretical Bioinformatics at the German Cancer Research Centre (DKFZ) in the group of Prof. Dr. Rainer König.

Heidelberg, November 2015

Ashwini Kumar Sharma

Dedicated to the loving memory of *Hazurbaba*

(1919 – 2003)

"The essence of all beautiful art, all great art, is gratitude"

— Friedrich Nietzsche

ACKNOWLEDGMENTS

I would like to take this opportunity to extend my heartfelt gratitude to all those individuals who have made my doctoral studies possible. Firstly, I would like to thank my mentor Prof. Dr. Rainer König who has been extremely supportive during my time in his lab. He has always encouraged me to pursue my ideas which have culminated into this thesis. I would also like to thank my thesis advisory committee members Prof. Dr. Peter Lichter, Prof. Dr. Christel Herold-Mende and Dr. Bernhard Radlwimmer whose critical suggestions have helped improved this work. I thank the members of my thesis defense committee for agreeing to evaluate my dissertation. I am very grateful to Prof. Dr. Roland Eils for his support and assistance during my doctoral period in his department.

I thank the Helmholtz International Graduate School for Cancer Research Fellowship and the Federal Ministry of Education and Research (BMBF), Germany for supporting my doctoral studies. I would like to acknowledge that the results of this study are mainly based on the publicly available genomic data generated on multiple cancer types by The Cancer Genome Atlas (TCGA) research network (`http://cancergenome.nih.gov/`).

I would like to sincerely thank the members of Network Modeling group and our extended group from Eilslabs. It has been a privilege to work with so many talented individuals who have rubbed off some of their own scientific brilliance onto me during my time here. I would like to extend my gratitude to my friends and colleagues – Anna Dieckmann, Lorenz Maier, Voker Ast, Dr. Moritz Aschoff, Dr. Tobias Bauer, Dr. Rosario M. Piro, Dr. Marcus Oswald, Dr. Zita Soons, Dr. Apichat Suratanee, Dr. Kitiporn Plaimas, Dr. Richa Batra and Alexandra Poos who have been tremendously helpful throughout this journey. In particular, I thank Anna and Lorenz who have helped me in the German translation of the section *Zusammenfassung* in this thesis and who in numerous occasions has helped me in both professional and personal capacities. I sincerely thank the members of the administrative and support staff in our department – Karl-Heinz Groß, Rolf Kabbe, Corinna Sprengart, Dr. Jan Eufinger, Ulrike Conrad, Dr. Julia Ritzerfeld and Manuela Schäfer, who work behind the scenes but have made so many aspects of my work easier for me every day. In addition, I thank Dr. Lindsay Murrells and Heike Langlotz from the DKFZ administration office for their extensive help over the years.

Outside of my department and work circle, I have had the great opportunity to know many people who have been a source of constant support and comfort during my time in Heidelberg. In particular, I would like to thank Uttara Ghosh for her unconditional support and faith in me during my research period. I sincerely thank all my dear friends who have been immensely helpful in providing the moral support during my doctoral thesis – Dr. Nisit Khandelwal, Dr. Ivana Dokic, Dr. Andrea Mairani, Dr. Taronish Dubash, Dr. Serena Bonifati, Francesco Marini, Dr. Devanjali Dutta, Dr. Mudita Pincha, Vignesh Natarajan, Shariq Ansari, Dr. Sushma Nayak, Dr. Sonia Savant, Dr. Kirti Shukla, Ria Thomas and Sita Priya Moorthi.

Finally, I would like to thank my family – my mother Dr. Raj Kumari Dahal, father Purna Kumar Sharma and sister Anjali Sharma for shaping me into what I am and towards whom I am eternally indebted.

PUBLICATIONS *(IN ORDER OF CONTRIBUTION)*

THESIS RELATED

1. **Sharma AK**, Eils R and König R. *Tumor metabolic reprogramming by copy number co- alterations of proximal enzyme coding and cancer causing genes*. (manuscript under review, 2015)

2. **Sharma AK** and König R. *Metabolic network modeling approaches for investigating the "hungry cancer"*. Seminars in Cancer Biology; Vol.23-4; 2013. (Review paper)

OTHER COLLABORATIVE CONTRIBUTIONS

1. Shukla K, **Sharma AK**, Ward A, Will R, Hielscher T, Balwierz A, Breunig C, Münstermann E, König R, Keklikoglou I and Wiemann S. *MicroRNA-30c-2-3p negatively regulates NF-kB signaling and cell cycle progression through downregulation of TRADD and CCNE1 in breast cancer.* Molecular Oncology; Vol.9-6; 2015.

2. Khandelwal N, Breinig M, Speck T, Michels T, Kreutzer C, Sorrentino A, **Sharma AK**, Umansky L, Conrad H, Poschke I, Offringa R, König R, Bernhard H, Machlenkin A, Boutros M and Beckhove P. *A high-throughput RNAi screen for detection of immune-checkpoint molecules that mediate tumor resistance to cytotoxic T lymphocytes.* EMBO Molecular Medicine; Vol.7-5; 2015.

3. Ummanni R, Mannsperger HA, Sonntag J, Oswald M, **Sharma AK**, König R, Korf U. *Evaluation of reverse phase protein array (RPPA)-based pathway-activation profiling in 84 non-small cell lung cancer (NSCLC) cell lines as platform for cancer proteomics and biomarker discovery.* Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics; Vol.1844-5; 2014.

MANUSCRIPT UNDER PREPARATION

1. Abdel-Haleem AM, **Sharma AK**, Piro R, Tönjes M, Plaimas K, Hamdy N, Eils R, Radlwimmer B, Lewis NE and König R. *Metabolic modeling reveals a differential requirement for PHGDH and BCAT1 in glioblastoma.* (in preparation, 2015)

2. Ansari SS, **Sharma AK,** Ivanova E, König R, Berger M. *Cell–cycle is affected by erufosine treatment in head and neck tumors.* (in preparation, 2015)

CONTENTS

LIST OF FIGURES

LIST OF TABLES

ZUSAMMENFASSUNG

HINTERGRUND

Krebs ist eine direkte Konsequenz von genomischen Aberrationen, wie somatische Kopienzahlveränderungen, die häufig im Krebsgenom auftreten und nicht nur onkogene Gene betreffen, sondern auch mehrere Passenger oder potentielle Ko–driver Gene. Eine intrinsische Eigenschaft, die aus einer solchen Störung des Genoms resultiert, ist die Deregulierung der metabolischen Landschaft des Tumors. Tatsächlich wurden mehrere metabolische Gene als Onkogene, Tumor-Suppressor-Gene oder als Ziele onkogener Signalwege identifiziert.

ERGEBNISSE

In dieser Arbeit legen wir dar, dass eine lineare Nähe metabolischer und krebserregender Gene im Genom zu metabolischen Umgestaltungen durch Ko–variationen der Kopienzahl führen kann. Wir haben beobachtet, dass Paare von Krebs unterstützenden metabolischen Genen unerwartet oft nah beieinander im Chromosom positioniert sind und sich in Bereichen mit veränderter Kopienzahl befinden, in denen sie über alle analysierten Krebsarten hinweg entweder besonders häufig ko-deletiert oder ko-amplifiziert sind. Wir haben die Analyse–Pipeline Identification of Metabolic Cancer Genes *(iMetCG)* entwickelt, um die funktionellen Auswirkungen solcher Ko–variationen auf den onkogenen Metabolismus herzuleiten und zu unterscheiden, welche Gene den Krebsmetabolismus tatsächlich antreiben, und welche neutral agieren. Zusätzlich beobachteten wir, dass die ermittelten metabolischen Krebsgene eine höhere Netzwerkkonnektivität haben, Indikatoren reduzierten Überlebens sind, und eine signifikanten Überlappung mit bekannten metabolischen Krebsgenen und deren Eigenschaften bezüglich Isoformdiversität und Selektionsdruck haben.

SCHLUSSFOLGERUNGEN

Diese Doktorarbeit liefert neue Einsichten in die funktionellen Mechanismen der metabolischen Regulierung aufgrund von Ko–amplifikationen und Ko–deletionen und deren Auswirkung auf die Veränderungen der metabolischen Landschaft. Unser krebsübergreifender, von Genomdaten getragener Ansatz deckt einen bislang unbekannten gen-erischen Mechanismus zur großflächigen metabolischen Umprogrammierung in Krebszellen auf, basierend auf linearer Gennähe, und identifiziert 119 neue metabolische Krebsgene, die wahrscheinlich an der Umgestaltung des Krebszell-Metabolismus beteiligt sind. Darüber hinaus werden unsere neu identifizierten metabolischen Krebsgene als eine wichtige Ressource der experimentellen tumormetabolischen und Genom–orientierten Forschung dienen und deren Geltungsbereich erweitern.

SUMMARY

BACKGROUND

Cancer is a direct consequence of genomic aberrations, such as somatic copy number alterations that frequently occur in the cancer genome affecting not only oncogenic genes, but also multiple passenger and potential co-driver genes. An intrinsic feature resulting from such a disruption of the genome is deregulation of the tumor metabolic landscape, as a result of which, multiple metabolic genes have been identified as oncogenes, tumor suppressor genes or targets of oncogenic signaling.

RESULTS

Here we elucidate that linear proximity of metabolic and cancer-causing genes in the genome can lead to metabolic remodeling through copy number co-alterations. We observed that cancer-metabolic gene pairs are unexpectedly often proximally positioned in the chromosomes and share loci with altered copy number, thus being either co-deleted or co-amplified across all cancers analyzed (19 cancer types from The Cancer Genome Atlas). We have developed an analysis pipeline – *I*dentification of *Met*abolic *C*ancer *G*enes (*iMetCG*), to infer the functional impact on oncogenic metabolism from such co-alteration events and delineate genes truly driving cancer metabolism from those that are neutral. Using this approach, we have identified novel and well known metabolic genes that target crucial pathways relevant for tumors. Moreover, using these identified metabolic genes we were able to classify tumors based on its tissue and developmental origins. We further observed that these putative metabolic cancer genes (identified across cancers) had higher network connectivity, were indicators for patient survival, had significant overlap with known cancer metabolic genes and shared similar features with known cancer genes in terms of their isoform diversity, evolutionary rate and selection pressure.

CONCLUSIONS

This thesis provides novel insights into the functional mechanism of metabolic regulation and rewiring of the metabolic landscape in cancer cells. Our pan-cancer, genomic data driven approach revealed a hitherto unknown generic mechanism for large scale metabolic reprogramming in cancer cells based on linear gene proximities and identified 119 new metabolic cancer genes likely to be involved in remodeling tumor cell metabolism. Furthermore, our newly identified metabolic cancer genes will serve as a vital resource to the experimental community engaged in tumor metabolism and genomics research to further expand the scope of this field.

1 SCOPE

1.1 AIMS

[1]The major objective of this thesis is to combine the concepts of gene order, cancer gene islands and cumulative gene dosage to investigate the resulting consequences for metabolic alterations in tumors. The linear arrangement of genes in chromosomes i.e. the gene order is now known to be non-random for the human genome, potentially due to functional and evolutionary constraints on genome organization. Furthermore, co-expressed gene clusters in the chromosomes are often under co-transcriptional control and these genes are simultaneously regulated by local changes in the chromatin state. Also, with the increasing availability of high throughput tumor genomics data, a major focus of oncogenomics research is on differentiating "driver" and "passenger" mutations. However, at the same time, a contrasting view of cancer gene islands and cumulative gene dosage are emerging concepts in cancer genomics. The large number of genes affected during carcinogenesis which was previously thought to be non-functional, a result of collateral damage from driving events is now believed to influence cancer survival and proliferation. In context of these concepts, we addressed the following issues in our study –

1. Is there any order in how metabolic and cancer genes are linearly positioned in the chromosomes? Are the corresponding gene-pairs relatively close or farther apart?
2. If there is an order, does it affect the occurrence frequency of copy number co-alterations events between metabolic and cancer genes?
3. Do co-altered metabolic genes have any functional impact on tumor metabolism or are they just neutral events?

1.2 BACKGROUND

Tumorigenesis can occur through numerous mechanisms giving the malignant cells ceaseless proliferative advantages (Hanahan and Weinberg, 2011). Clonal evolutionary processes result

1 Parts of this chapter have been taken from (Sharma et al., 2015)

in a perpetual conflict between the fitness of the host cell community and cancer initiating cells. The latter are governed by an intrinsic driving force towards uncontrolled proliferation, and if sustained, eventually lead to a neoplastic state. Somatic evolutionary forces could have shaped inherent vulnerabilities in the host genome, thus providing a cancer initiating potential, a consequence of various genetic trade-offs during the selection process. These trade-offs are reflected by changes to the fitness balance that determine tumor progression or expulsion via protumorigenic or tumor suppressive mechanisms. The existence of tumor suppressors and oncogenes in the genome makes this evolutionary dilemma apparent (Merlo et al., 2006; Greaves and Maley, 2012).

Several studies have shown that gene order is not random in eukaryotes, but is subjected to natural selection driven by functionality (Hurst et al., 2004; Singer et al., 2005; Semon and Duret, 2006). Co-expressed genes, in particular housekeeping genes, cluster together in the chromosomes, and are often from the same biochemical pathways (Cohen et al., 2000; Caron et al., 2001; Lercher et al., 2002; Lee and Sonnhammer, 2003). Accordingly, these gene clusters can be co-regulated (Gierman et al., 2007; Purmann et al., 2007; Davila Lopez et al., 2010) and genomic regions with a higher density of cancer genes could be under co-transcriptional control (Glinsky et al., 2003).

Somatic copy number alterations (SCNA) in cancer cells can affect large segments of the genome and these susceptible regions often harbor genes affecting cellular proliferation. Typically, genes occupying these "cancer gene islands" were previously thought to be bystanders and collaterally affected from SCNA targeting a cancer-specific gene. Interestingly, these co-affected genes are now considered to contribute towards cancer potency via "cumulative haploinsufficiencies" (Solimini et al., 2012; Davoli et al., 2013). We would lose information on these secondary effectors when focusing only on the major cancer driving gene affected by such high frequency SCNA events. However, a complete cellular process might be perturbed from a global standpoint. Recent studies have also reported specific SCNA-driven functional co-alterations between proximal cancer and metabolic genes in renal cell carcinomas, glioblastomas and breast cancers (Bashashati et al., 2012; Gatto et al., 2014). Moreover, co-deletion of *MTAP* with the *CDKN2A/2B* tumor suppressor gene is a long known example that occurs across multiple cancer types affecting the biochemical process of adenosine and methionine salvage (Carson et al., 1988; Bertino et al., 2014).

In this study we have systematically explored the metabolic consequences of linear proximity between metabolic and cancer genes and their concurrent copy number co-alterations across a wide range of cancer types. Based on our incipient understanding that altered metabolism is crucial to carcinogenesis (Kroemer and Pouyssegur, 2008; Schulze and Harris, 2012; Sharma and König, 2013), we assessed the functional relevance of such events and elucidated their potential oncogenic effects and the resulting impact on cancer metabolic remodeling.

1.3 MAJOR FINDINGS AND RELEVANCE

In order to answer these pertinent questions, we have developed an analysis pipeline that primarily employs computational oncogenomic analysis using robust statistical methods across 19 cancer types from The Cancer Genome Atlas (TCGA) for the identification of a generic yet novel mechanism causing reprogramming of cancer metabolism. The major findings of this work are –

1. Cancer causing (oncogenes or tumor suppressors) and metabolic genes are significantly closer together in the genome than cancer-nonmetabolic gene pairs.
2. This linear proximity is exploited through increased propensity for copy number co-alterations of such gene pairs across cancer types.
3. A computational pipeline, *iMetCG*, was developed that integrates copy number co-alteration and co-expression information for the identification of significantly altered and proximal cancer-metabolic gene pairs.
4. These pairs are further prioritized for the discovery of novel cancer metabolic genes using *a priori* defined functionally relevant gene sets (like essential and rate-limiting genes) and identifying those co-altered metabolic genes whose network neighbors are also perturbed.
5. Co-altered metabolic genes identified using our pipeline (i.e. a 113 metabolic gene signature from 16 cancers types) enables tumor classification in tune to its tissue and developmental origins.
6. The identified cancer metabolic genes reprogram cancer metabolism by targeting well known cancer metabolic pathways like nucleotide, lipid, amino acid and energy metabolism.

7. The evolutionary (*dN/dS* ratio, evolutionary rate, isoform diversity and gene connectivity) and phenotypic properties (cancer survival and enrichment in known cancer metabolic genes) of these novel metabolic cancer genes are very similar to known cancer genes, thus supporting our hypothesis that they are potentially metabolism specific oncogenic drivers.

This thesis provides novel insights into the functional mechanism of metabolic regulation in cancer cells. It highlights a hitherto unknown generic mechanism for rewiring of tumor metabolic landscape based on linear gene proximities between cancer causing and metabolism specific genes in the chromosomes. This linear proximity translates into their higher rates of copy number co-alterations leading to large scale metabolic reprogramming in tumors. In this work, through integrative genomics based approaches and our newly developed statistical analysis pipeline *iMetCG*, we have identified 119 putative metabolic cancer genes likely to be responsible for remodeling cancer cell metabolism. Moreover, these identified metabolic cancer genes will serve as a vital resource to the experimental community engaged in tumor metabolism research.

1.4 OUTLINE OF THE THESIS

In this thesis, the major focus is on the effect of genomic alterations, specifically the effect of SCNA on tumor metabolism which will be presented in a typical cause and consequence study format. The main motivation of this present work is to infer causal genomic mechanisms leading to changes in the tumor metabolic landscape using computational approaches. In the introductory chapter (Chapter 2), three major concepts namely – rewiring of cancer metabolism, gene order and effect of cumulative gene dosage on carcinogenesis potential will be discussed in detail. Linking these ideas form the backbone of this study. In the next chapter (Chapter 3), the methodological aspects of this work will be elaborated with regards to integration of various datasets and development of a novel statistical pipeline to address the questions posed in this thesis. This will be followed by the presentation of results and discussions (Chapter 4), here the findings of this thesis in context of our current knowledge will be analyzed. Finally, in the concluding chapter (Chapter 5), a brief summary of the entire work will be presented highlighting the novelty and relevance of this study along with suggestions for future studies.

2 INTRODUCTION

2.1 CANCER - GENETIC CAUSES AND CONSEQUENCES

[1]Cancer is one of the leading causes of mortality and morbidity worldwide with an estimated 14.1 million new cancer related cases, 8.2 million cancer caused deaths and 32.6 million people living with cancer (diagnosed within the last 5 years) as of 2012 (Ferlay et al., 2015). To put these numbers into perspective 1 out of 2 individuals in Germany, United Kingdom or the United States will be diagnosed with cancer in their lifetime (Kaatsch et al., 2014; Hayat et al., 2007)(Cancer Research UK, 2015). The most common types of cancers diagnosed worldwide (in both males and females) are of the lung, breast, prostate, cervix, colon, stomach and liver (Ferlay et al., 2015). A common misconception regarding cancer is the perception that it is a disease of modern times, however, cancer has always ocuured during the course of human history. The increased incidence of cancer seen in recent times is mainly attributed to vastly improved healthcare resulting in increased life expectancies. In fact, the earliest description of cancer dates back to almost 3000 BC recorded in an Egyptian textbook describing a surgical method for breast cancer while strong physical evidence in the form of fossilized records suggestive of osteosarcoma has been discovered from Egyptian mummies. Hippocrates (460–370 BC), considered as the father of medicine, had already witnessed and studied tumors in ancient Greece and is credited to have coined the word *carcinos* and *carcinoma* while describing tumorous growth in his patients (Mukherjee, 2010).

1 out of 2 people will be diagnosed with cancer in their lifetime

More than a century ago, Theodor Boveri, suggested that cancer might be a result of chromosomal alterations that leads to uncontrolled cell division (Boveri, 2008). After decades of extensive cancer research, it is now firmly established that cancer is indeed a direct consequence of cumulative accumulation of multiple genomic aberrations like mutations, somatic copy number changes (SCNA) and translocations (Vogelstein et al., 2013). Point mutations result in single nucleotide changes which can be synonymous (i.e. silent or neutral) causing no change in the type of amino acid being translated due to inherent codon redundancy. Mutations can also be non-synonymous

1 Parts of this chapter has been taken from (Sharma and König, 2013)

(i.e. missense, nonsense and frame-shift mutations) where there is a change in the codon resulting in a concordant change in the translated amino acid. This however may or may not affect protein function based on factors like physiochemical similarities between the original and changed amino acids and the position of amino acid change in the protein structure. For example, a change of an amino acid at the protein loop regions may have less severe effect on protein function when compared to a change in the active site of a protein. In more extreme cases, point mutations may lead to conversion of an amino acid coding codon into a stop codon (nonsense mutation) which would result in premature termination of protein translation. Similarly, a single nucleotide insertion/deletion (indel) in the genome can cause a complete change in the sequence of amino acids being translated, thus creating a non-functional protein. SCNA is the most common type of genomic aberration in tumors that occurs due to a defective replication machinery leading to deviations from the diploid number of chromosomes (aneuploidy) or focal addition/removal of DNA material (amplifications or deletions). SCNA has a severe impact on the genomic blueprint because it can simultaneously affect multiple genes causing a drastic change in gene dosage. Such changes may affect the entire chromosome, chromosomal arm or focally on relatively smaller segments of the chromosome. Furthermore, chromosomal alterations may result in fusion genes due to translocation events (balanced, unbalanced or centromeric) where segments of the chromosome are exchanged between the same or different chromosome (Fig.1A) (Griffiths et al., 2010).

Cancer is a consequence of progressive accumulation in genomic aberrations

These genetic assaults sustained by a normal cell leads to its neoplastic transformation which is manifested by increasing cellular division and growth due to sustained growth signals, inhibition of growth suppressors, replicative immortality, genome instability, inflammation, angiogenesis, metastasis, evasion from apoptosis, circumvention of immune-surveillance and metabolic reprogramming (Hanahan and Weinberg, 2011) (Fig. 1B). Among the different hallmarks typically exhibited by tumors, studies on metabolic alterations in cancer cells is fast gaining traction due to observations that in tumors, many metabolic genes are mutated or are downstream targets of deregulated signaling (Kroemer and Pouyssegur, 2008; Cairns et al., 2011; Schulze and Harris, 2012). Almost all of the "cancer hallmarks" are involved in crosstalk and metabolism seems to be uniquely placed at the crossroads. For example, increased availability of growth signals or resistance to growth suppressors leads to rapid cell proliferation. This typically involves high

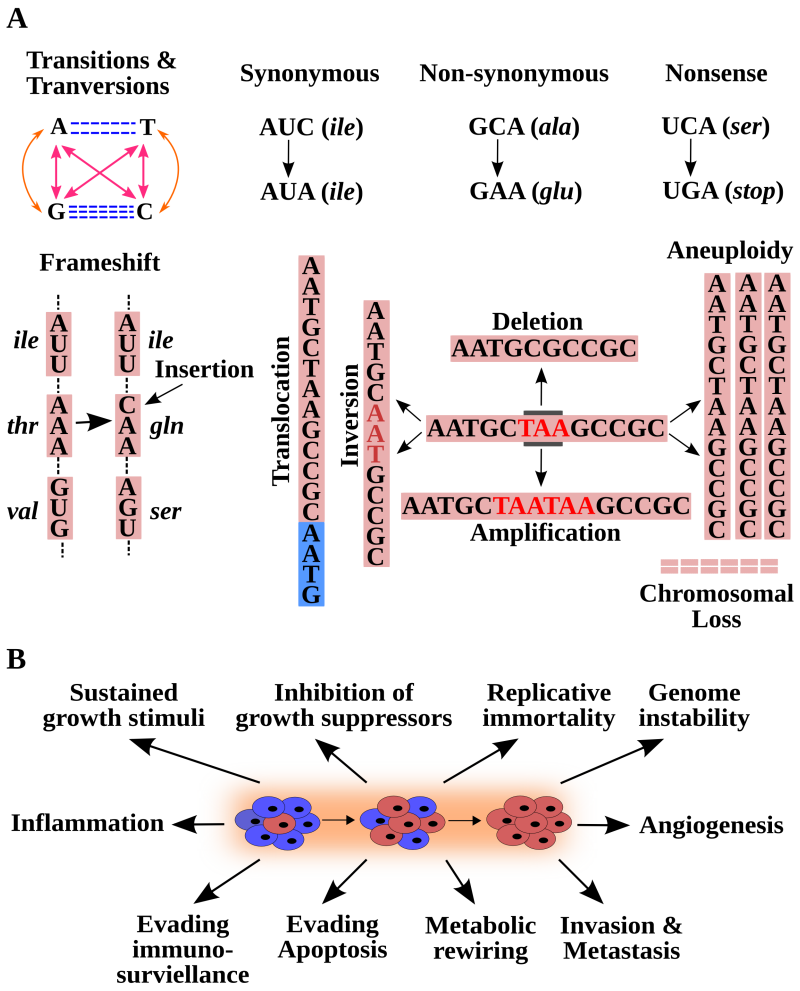


Figure 1: **Cancer - causes and consequences.** (**A**) Different types of genomic aberrations that facilitate the emergence of the cancer phenotype. (**B**) The typical "hallmarks of cancer"(Hanahan and Weinberg, 2011) exhibited by tumors.

consumption of glucose and glutamine for high cellular biomass generation. Similarly, metabolic reprogramming is involved in metastasis where peripheral membrane proteins maintaining the spatial fidelity of cells in context of its tissue organization are altered leading to cellular detachment from the extracellular matrix. These membrane proteins are post-translationally modified with complex metabolic products composed of fatty acids and sugars like glycosylphosphatidylinositol, alterations in which promotes cell invasion (Pinho and Reis, 2015). Furthermore, the tumor microenvironment too plays a crucial role in tumorigenesis where it has been elegantly shown using a two compartment model that cancer and cancer associated stromal cells are involved in metabolite exchange to optimize nutrient utilization. Lactate generated in glycolytic stromal cells are taken up by the cancer cells to be utilized for energy gener-

Deregulation of the tumor metabolic landscape is an "emerging hallmark of cancer"

ation – a process called the "reverse Warburg effect" (Pavlides et al., 2009). All these observations strongly support the fact that metabolic alterations are indeed a fundamental requirement for tumorigenesis and is extensively involved in supporting the various so called non-metabolic functions of tumor cells. In the following sections the crucial role of cancer metabolism in tumorigenesis will be elucidated in detail.

2.2 THE STRUCTURE AND FUNCTION OF HUMAN METABOLISM

Metabolism is the collective of thousands of chemical reactions occurring at a given condition resulting from the intricate interplay of various regulatory processes. It is among the most basic functional phenotype of the cell and is shown to be highly conserved across species. The study of these dynamic metabolic processes and their capabilities helps to identify the fundamental properties of living systems. Metabolism can be considered as a thermodynamic open-system in which source substrates of high value are being processed through a well established and interconnected biochemical conversion system, strictly obeying physiochemical principles, generating useful intermediates and finally resulting in the release of byproducts (Sharma and König, 2013). In this section, the daunting complexity of the metabolic network will be broken down into simpler recurring processes and the rules by which metabolism is controlled will be explained highlighting that there is an inherent order in how metabolic processes work.

The complexity of human metabolic network can be broken down into small number of recurring modules

The human metabolic network has around 7440 chemical reactions consisting of 5063 metabolites catalyzed by enzymes translated from 1789 genes (Thiele et al., 2013). However, this complex dimensionality of the metabolic network can be vastly reduced by categorizing all of these reactions and metabolites into 99 well defined subsystems/pathways and classifying enzymatic reactions into 6 major types of chemical transformation process that they catalyze (Table. 1) (Thiele et al., 2013; Cornish-Bowden, 2014). Metabolism can be further represented by a bipartite graph connecting each pair of reactions by a metabolite which is the substrate of one and the product of the other reaction. Such a representation facilitates analysis of the topological properties for the metabolic network. Moreover, metabolic networks like any other scale-free network in biological systems closely follows the power law as a result of which very few metabolites are connected to most of the reactions in the network (like pyruvate, glutamine, ATP, CoA *etc*). These highly connected metabolites are called hubs while the rest of the

Table 1: **Major classes of metabolic enzymes and its catalytic function†.**

EC class	Name	Function	Example
1	Oxido-reductases	Oxidation-reduction,electron transfer	Lactate dehydrogenase (*LDH*)
2	Transferase	Chemical group transfer	Alanine amino-transferase (*GPT*)
3	Hydrolases	Transfer of a chemical group to water, breakage of bonds to form two fragments	Esterase D (*ESD*)
4	Lyases	Non hydrolytic addition or removal of chemical groups, double bonds formed	Fumarate hydratase (*FH*)
5	Isomerases	Isomerization reactions, rearrangement of atoms in a molecule	Triose phosphate isomerase (*TPI*)
6	Ligases or synthetase	Joining of two molecules with ATP (or any energy rich) hydrolysis	Long-chain-fatty-acid-CoA ligase 1 (*ACSL1*)

†also see – `http://www.chem.qmul.ac.uk/iubmb/enzyme/`.

metabolites are connected to only handful of reactions. Also, it has been observed that the network diameter is relatively small which means that any two metabolites in the network can be connected in few steps along the shortest path (Jeong et al., 2000). In context of function, the metabolic system processes 4 major biomolecules required for cellular sustenance, which are – carbohydrates, lipids, nucleotides and amino acids. Almost all metabolic reactions driven to generate and transform these metabolites can be broadly classified into catabolic and anabolic processes where the substrates are either broken down to generate energy or the energy generated is used to synthesize new products respectively. The complexity of metabolic networks can be further decomposed into smaller functional units be-

cause many reactions are related and they converge into crucial metabolic hubs, similarly many reactions also diverge from a common precursor molecule to generate a wide variety of related metabolites (Berg et al., 2002; Nelson et al., 2008).

In 1930's Hans Krebs first described the three steps of catabolism which involves sequentially degradative processes converging into energy production through oxidation of consumed food. Furthermore, these degradative processes are tightly coupled to anabolic reactions and depending on the metabolic requirements, an optimal path is chosen by the cell while maintaining the intricate balance between these complementary processes.

Acetyl-CoA is a crucial metabolite connecting anabolic and catabolic processes

In the first phase, all carbohydrates, proteins and fats from the food source are degraded into their basic building blocks namely the hexose sugars, amino acids and glycerol/fatty acids, this is the preparatory phase and no energy is produced. In the second phase, metabolic outputs of the previous stage undergo chemical transformation through a series of reactions to finally converge into acetyl-CoA generation. Carbohydrates are generally consumed in the form of (mono/di/poly)–saccharides (like fructose, lactose, sucrose or starch) which are composed of either aldoses (like glucose, mannose or galactose) or ketoses (like fructose), in both cases the sugars converge into the glycolysis pathway where the sugar carbon backbone is oxidized to generate pyruvate which is then further converted to acetyl-CoA that goes into the tricarboxylic acid (TCA) cycle, also known as the Krebs cycle (Fig.2) .

Proteins are first degraded into its constituents i.e. 20 possible amino acids by proteases. These amino acids can be salvaged and used for anabolic purposes. However, in the case of energy requirement, they are further degraded to generate pyruvate (which is further converted to acetyl-CoA) or intermediates of the TCA cycle (like α-ketoglutarate or oxaloacetate). In all cases of amino acid degradation the downstream products converge into the TCA cycle (Fig.2).

Dietary fats are stored in the adipose tissues as energy reserves. During energy starvation these fatty acids are coupled with coenzyme A (CoA) to generate fatty acyl-CoA which is then transported into the mitochondria where through the process of β-oxidation, either acetyl-CoA or succinyl-CoA is generated depending on the number of backbone carbons, both of which converge into the TCA cycle. In the final stage, acetyl-CoA is completely oxidized through the Krebs cycle to generate electrons which is then transferred to reduced electron carriers like (NAD+, FAD), this process creates a proton gradient which is used to generate energy in the form of ATP during the oxida-

tive phosphorylation process (OXPHOS) (Fig.2). Thus, despite the high complexity of metabolic reactions there exist few basic marcromolecules which undergo large number of diverse chemical transformation steps to yield few key metabolic precursors, all of which are processed through the same energy generating processes. Similarly, in anabolic reactions few crucial metabo-

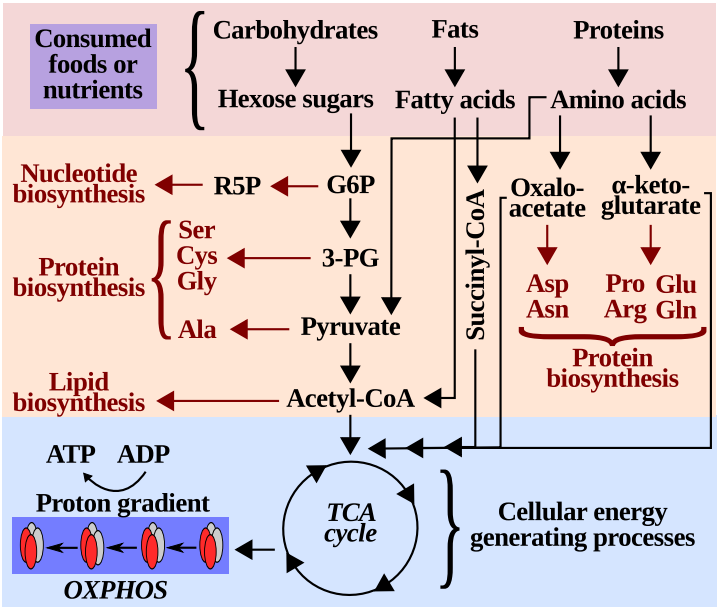


Figure 2: **Three phases of food metabolism.** Consumed food in the form of carbohydrates, proteins or fats are initially degraded into its constituent chemical units (pink panel). These endproducts are further transformed into few crucial metabolic precursors that enter into the TCA cycle (yellow panel). Finally, oxidative phosphorylation (OXPHOS) generates the bulk of cellular energy by harvesting the chemical proton gradient into ATP. The catabolic processes (shown in black arrows) are tightly coupled with biosynthetic reactions (shown in red arrows) that generate cellular biomass. G6P - Glucose-6-phosphate, R5P - Ribose-5-phosphate, 3-PG - 3-phosphoglycerate, Ser - Serine, Gly - Glycine, Cys - Cysteine, Ala - Alanine, Asp - Aspartate, Asn - Aspargine, Pro - Proline, Arg - Arginine, Glu - Glutamate and Gln - Glutamine.

lites through successive and connected chemical reactions can generate the vast variety of biosynthetic end products. For example, glycolytic intermediates (like 3-phosphoglycerate and pyruvate) and TCA cycle intermediates (like α-ketoglutarate and oxaloacetate) can generate all of the non-essential amino acids in the cell which is then used for protein biosynthesis. Also, in nucleotide metabolism a similar pyramidal structure for coupled chemical reactions exists where ribose-5-phosphate, glutamine and aspartate can produce the entire set of purines and pyrimidines (mono/di/tri–)phosphates used for DNA synthesis in the cell (Berg et al., 2002; Nelson et al., 2008).

In recent years there has been a reemergence of interest in tumor metabolism research

The high redundancy and connectivity exhibited by metabolic networks makes its highly robust to external perturbations and facilitates various rewiring mechanisms to maintain its optimal efficiency. Such remodeling mechanisms and their functional consequences in a highly deregulated genomic state like tumors will be described in detail in the following sections.

2.3 CANCER METABOLISM

Rapidly dividing cells have have high biomass production

Rapidly proliferating cells exhibit distinct metabolic features to support their growth demands which typically manifest by increasing rates of energy metabolism to support lipid, nucleotide and amino acid biosynthesis. Moreover, the catabolic as well as anabolic processes within the cell are uniquely balanced to augment these growth requirements. Utilizing the vast redundancy and connectivity of the metabolic network, cells can rewire metabolic routes to optimize its growth potential and increase biomass (DeBerardinis et al., 2008; Kroemer and Pouyssegur, 2008; Dang, 2012; Schulze and Harris, 2012). This metabolic remodeling of the tumor cell is tightly regulated by its microenvironment, cellular physiology and various (epi)genetic mechanisms (Fig.3) (Cairns et al., 2011; Ward and Thompson, 2012). Detailed understanding of this reprogramming of cellu-

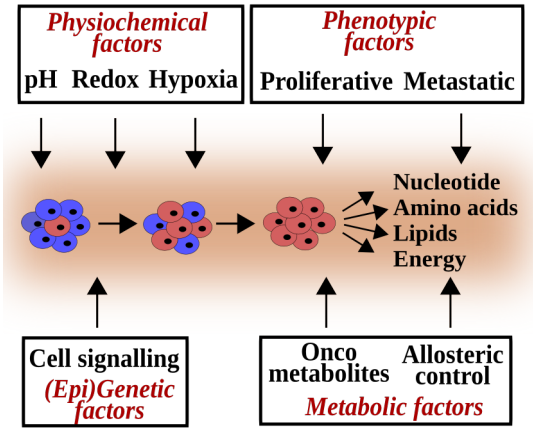


Figure 3: **Tumor metabolism.** Rapidly proliferating cells have high energy demands to support the biosynthesis of cellular biomass consisting mainly of nucleotides, lipid and proteins. These metabolic processes are tightly regulated within the tumor cell by the microenvironment, physiology and signaling mechanisms.

lar bioenergetics is crucial in the context of cancer as it can highlight critical links that fuels neoplastic growth which can be exploited to disrupt the diseased state using therapeutic interventions (Galluzzi et al., 2013). Besides the known and well-studied

causalities for tumorigenesis, conceptual progress within the last decade has established that reprogramming of energy metabolism is an "emerging hallmark of cancer" (Fig.1B) which is now regarded indispensable for tumor formation (Hanahan and Weinberg, 2011; Sharma and König, 2013). Furthermore, it is increasingly becoming clear that changes in the tumor metabolic landscape leads to rapid generation of reactive oxygen species (ROS). This causes an enhanced genomic mutation load, thus implicating altered metabolism as a direct cause for genomic aberrations (Cooke et al., 2003; Cairns et al., 2011). Over the last decade there has been a resurgence of active research in the field of cancer metabolism that has rightfully placed metabolism back into the center stage of cancer research. This is highlighted by an exponential growth in the number of scientific publications in this field over the last few years (Fig.4). This reemergence of scientific interest is mainly because recent

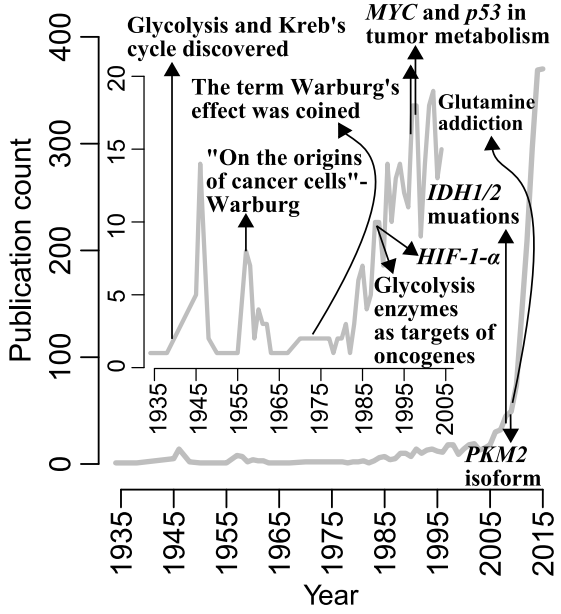


Figure 4: **The reemergence of tumor metabolism research.** A conservative search for the phrase "cancer/tumor metabolism" or "metabolic rewiring/reprogramming" was carried out in Pubmed (July, 2015) to get the year wise counts for publications containing this phrase. Crucial events that have defined the evolution of cancer metabolism research are highlighted (also see(Koppenol et al., 2011)).

genome scale analysis of cancers and its high-throughput functional studies have revealed that metabolic genes are crucial oncogenes or tumor suppressor genes and are also major downstream targets of deregulated signaling programs in tumors. This deregulation results in rewiring of the metabolic circuitry conferring an exploitative metabolic advantage for the tumor

cells which carries a distinct benefit for tumor survival and lays the basis for its unbound progression. Hence, understanding the aberrant mechanism of tumor metabolism is of central interest in biomedical research (Sharma and König, 2013).

2.3.1 *Alterations in central carbon metabolism pathways*

Alterations in the central carbon metabolism supports the enhanced metabolic requirements in tumor cells

The central carbon metabolism consists of a set of interconnected metabolic subsytems (glycolysis, pentose phosphate pathways and Kreb's cycle) that are involved in energy and biomass production. Glycolysis involves the oxidative breakdown of six carbon sugars (glucose) into three carbon products (two pyruvates) in ten successive chemical transformation steps, with the generation of energy (2 molecules of ATP) and reducing equivalents (2 molecules of NADH) (Nelson et al., 2008). Almost, every enzyme involved in glycolysis is targeted by deregulated signaling processes in tumors (see the section on regulation of tumor metabolism below), starting from the initial glucose transporters (*GLUT*s, that supply glucose for glycolytic processing) to lactate/pyruvate transporters (*MCT1* and *MPC1*, that funnel glycolysis end products outside the cell or into the TCA cycle respectively) (Macheda et al., 2005; Kennedy and Dewhirst, 2010; Schell et al., 2014). Among others, enzymes deregulated in tumors include all of the major glycolytic flux controlling enzymes like hexokinases *(HK1/2)* (Wolf et al., 2011), pyruvate kinases *(PKM2)* (Vander Heiden et al., 2010), lactate dehydrogensaes *(LDHA)* (Le et al., 2010) and phosphofructokinases *(PFK1*) (Mor et al., 2011) which are also responsible for generating crucial intermediates for anabolic processes.

Intermediates of glycolysis are funneled into branching pathways for the generation of other essential metabolites. For example, glucose-6-phosphate enters into the oxidative arm of the pentose phosphate pathway (PPP) while fructose-6-phosphate and glyceraldehyde-3-phosphate enter into the non-oxidative arm of PPP, both leading to the production of ribose-5-phosphate which is an essential precursor for nucleotide biosynthesis. Additionally, the oxidative arm also generates NADPH which is used in many anabolic reactions (Berg et al., 2002). In tumors, multiple enzymes involved in PPP are deregulated, for example, the NADPH generating enzymes – glucose-6-phosphate dehydrogenase and 6-phosphogluconate dehydrogenase are crucial for tumorigenesis and under tight oncogenic signalling control. Similarly, transketolases and transaldolases are crucial enzymes in the non-oxidative arm of PPP and their ex-

pression is elevated in multiple tumor types and is associated with poor patient survival (Patra and Hay, 2014).

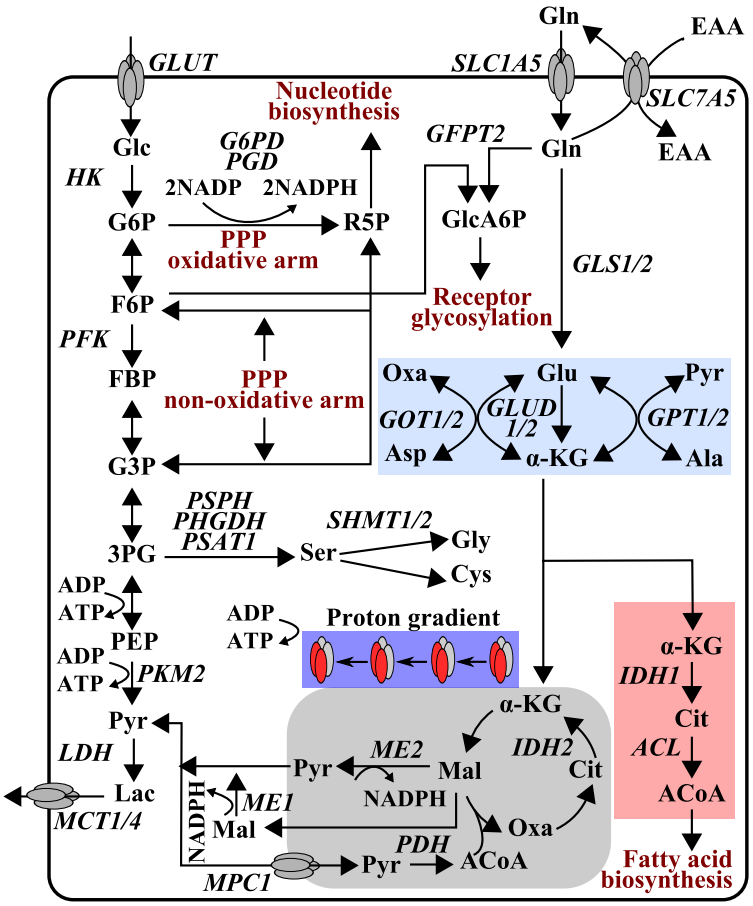


Figure 5: **The role of central carbon metabolism in tumors.** The central carbon metabolism pathways (glycolysis, pentose phosphate pathway and the TCA cycle) are highly altered in tumors. The commonly targeted metabolic enzymes in tumors and those that generate ATP and NADPH are shown in italics. Important metabolite transporters are shown in gray ovals. Glutaminolysis is shown within the blue shaded region, reductive carboxylation is shown within the pink shaded region and the mitochondrial compartment is highlighted in the gray shaded region. Glc - Glucose, G6P - Glucose-6-phosphate, R5P - Ribose-5-phosphate, F6P - Fructose-6-phosphate, FBP - Fructose bisphosphate, G3P - Glyceraldehyde-3-phosphate, 3PG - 3-phosphoglycerate, Ser - Serine, Gly - Glycine, Cys - Cysteine, PEP - Phosphoenolpyruvate, Pyr - Pyruvate, Lac - Lactate, ACoA - Acetyl-CoA, Oxa - Oxaloacetate, Cit - Citrate, α-KG - α-ketoglutarate, Mal - Malate, Gln - Glutamine, Glu - Glutamate, Asp - Aspartate, Ala - Alanine, GlcA6P - Glucosamine-6-phosphate, EAA - Essential amino acids.

Dihydroxyacetone phosphate (DAP), another intermediate of glycolyss is utilized for the generation of phosphatidic acid (PA), the backbone for all glycero(phospho)lipids. In tumors, it has been observed that enzymes involved in the conversion of DAP to PA are highly expressed and that phosphatidic acid is a crucial metabolic regulator of mTOR, which in turn regulates

diverse metabolic functions in cancer cells (Athenstaedt and Daum, 1999; Foster, 2007). Similarly, other glycolysis products like pyruvate and 3-phosphoglycerate are precursors for non-essential amino acid biosynthesis(Nelson et al., 2008). Apart from glycolysis, pyruvate can also be generated by the malic enzymes using malate from the TCA cycle, this process generates NADPH which is an essential reductant currency for tumors (Jiang et al., 2013). Furthermore, pyruvate can be converted into alanine which is used for protein biosynthesis and is also heavily secreted by the cancer cells, the reason for which is not clearly established (Ahn and Metallo, 2015). In breast cancers, it has been observed that 3-phosphoglycerate from glycolysis is utilized for serine and glycine production through the amplified *PHGDH* enzyme, which can be further utilized in nucleotide biosynthesis and one-carbon metabolic pathways (Possemato et al., 2011; Locasale et al., 2011). Also, the glycolytic intermediate, fructose-6-phosphate feeds into the hexosami-ne (amino sugars) biosynthetic pathway whose major function involves protein glycosylation and lipid modification. The aminotransferase enzyme *GFPT2*, utilizes glutamine and fructose-6-phosphate for the generation of the amino sugar, glucosamine-6-phosphate, which is further used in N-/O- linked glycosylation based modification of crucial membrane bound signaling receptors, the disruption of which perturbs tumor growth (Guillaumond et al., 2013).

Metabolic remodeling of the central carbon metabolism supports both anabolic and catabolic requirements of the cancer cell

As described above, the end product of glycolysis, pyruvate, enters into the Kreb's cycle for its complete oxidization into CO_2. This process generates 8 NADH, 2 $FADH_2$, 2 ATP and a proton gradient which is used to further harvest ATP during oxidative phosphorylation (OXPHOS) as a result of which there is net gain of 36 ATP upon complete oxidization of glucose (Nelson et al., 2008). Furthermore, the TCA cycle intermediates too contribute towards major biosynthetic processes like nucleotide, lipid, heme and amino acid biosynthesis (Fig. 5). Alterations in TCA cycle enzymes, their role and regulation in tumor metabolism is explained in the following sections.

In total, pathways involved in central carbon metabolism form the a fundamental unit of the human metabolic network as they are involved in supplying essential precursors and energy to the rest of the metabolic processes. Hence, the central carbon/energy metabolism is tightly regulated and is often perturbed during tumorigenesis causing a drastic shift in metabolic requirements compared to a normal cell.

2.3.2 *The Warburg effect and glutaminolysis in cancers*

The metabolic phenotype typically exhibited by most tumor cells involves elevated levels of glucose and glutamine consumption. These crucial metabolites serve as the bulk of carbon and nitrogen sources in tumors, expectedly, pathways involved in the utilization of these metabolites are among the most deregulated in cancer and under the tight control of oncogenic signaling (Dang, 2012).

Pioneering investigations by Otto Warburg in the 1920s revealed that cancer cells preferentially oxidize glucose through a fermentative glycolytic process even in the presence of sufficient oxygen. This metabolic phenotype shown by most cancer cells is called the Warburg effect which results in high levels of cellular lactate secretion (Warburg et al., 1956). The cause for this unique phenotype has been widely debated and several plausible mechanisms have been suggested. However, one of the initially proposed mechanism by Warburg himself suggesting a non-functional mitochondria in cancer cells has been ruled out (Koppenol et al., 2011). In fact, since then it has been shown that in many cancer cells OXPHOS actively generates ATP (Jose et al., 2011; Tan et al., 2015). Furthermore, it has been shown that cancer cells exhibit about 10 times higher glucose uptake and that for every 13 molecules of glucose consumed, 1 molecule of glucose goes through complete oxidation via OXPHOS while the remaining 12 glucose molecules undergo fermentative glycolysis producing lactate. Also, glucose oxidation in glycolysis and OXPHOS produces 2 and 36 molecules of ATP (18 times higher ATP in OXPHOS) respectively. However, with this reaction stoichiometry, in cancer cells with high glucose uptake, glycolysis alone can generate 24 molecules of ATP which is 2/3 of OXPHOS dependent ATP production (Koppenol and Bounds, 2009). It has also been suggested that in rapidly dividing cancer cells, the Warburg effect optimizes proliferation by generating high biomass for cellular maintenance leading to the Warburg phenotype. Furthermore, glycolytic dependency shown by tumors is also attributed to the fact that intermediate metabolites of this pathway serve as crucial precursors for macromolecular biosynthesis (Fig. 5) (Vander Heiden et al., 2009; Shlomi et al., 2011). Moreover, from an evolutionary viewpoint, it has been suggested that cancer cells adapt to the glycolytic phenotype because of hypoxic tumor microenvironment as a result of which these cells have to tolerate acidic conditions, those cancer cells surviving such harsh conditions are highly

The Warburg effect or aerobic glycolysis is a typical metabolic phenotype shown by most tumor cells

aggressive and resistant leading to its increased proliferation (Archetti, 2014).

Glutaminolysis is the conversion of glutamine into α-ketoglutarate via glutamate by glutaminase (*GLS1/2*), glutamate dehydrogenase (*GLUD1/2*) or aminotransferases (*GOT1/2*, *GPT1/2* etc). Glutamine is an important source of carbon and nitrogen in tumor cells where it is used extensively for replenishing nucleotides, lipids and non-essential amino acids (Fig. 5). Furthermore, a major role of glutamine is to maintain the redox balance in cancer cells by generating reducing equivalents (like NADH, NADPH) and glutathione pools (DeBerardinis and Cheng, 2010). Enzymes involved in glutaminolysis are deregulated in tumors and are targets of Myc and Kras oncogene dependent signaling (Wise et al., 2008; Son et al., 2013). Glutaminolysis is affected among others, in glioblastomas (DeBerardinis et al., 2007), breast cancer (Korangath et al., 2015) and pancreatic cancer (Son et al., 2013), all of which exhibit an increased dependency on glutamine for cellular metabolism, such that disruption of active glutaminolytic enzymes severely decreases tumor proliferation and is described as the "glutamine addiction" phenotype.

2.3.3 *Alterations in lipid metabolism*

In this section, various enzymes in lipid and fatty acid metabolism that show altered behavior in tumors are described, furthermore the metabolic routes rewired to support the biosynthetic demands are explained, highlighting how specific targeting of these changes makes the cancer cell vulnerable, suggesting the essential role of such metabolic reprogramming events.

Fatty acids are crucial building blocks within the context of cellular physiology and are involved in, among others – cellular structure maintenance, balancing energy homeostasis, signaling and anchorage for protein complexes. These cellular processes are disrupted in tumorigenesis and expectedly it has been shown that alterations in lipid metabolism through deregulation of specific enzymes involved in both fatty acid degradation and biosynthesis are essential features exhibited by rapidly dividing cells. Within cells, fatty acid requirements are satisfied by both endogenous (*de-novo* synthesis) or exogenous (uptake of food/nutrients via the bloodstream or from adipose tissues) nutrient pools through overlapping metabolic routes (Santos and Schulze, 2012).

Citrate is a crucial precursor metabolite in fatty acid biosynthesis and its sub-cellular localization has important implications on its metabolic fate. For instance, mitochondrial citrate

feeds into the TCA cycle while cytosolic citrate is routed through fatty acid biosynthesis via acetyl-CoA. Furthermore, there is active shuttling of citrate molecules between the mitochondrial and cytoplasmic compartments through the citrate transporting proteins (CTP), also known as citrate carriers (CiC), these proteins mainly belong to the *SLC25* family of transporters like the *SLC25A1* (Fig. 6). This transport process involves electroneutral exchange of di(tri)carboxylate molecules like citrate or isocitrate for malate, succinate or phosphoenolpyruvate (Mycielska et al., 2009; Gnoni et al., 2009). Given the essentiality of fatty acid biosynthesis for tumor survival/progression and the central role of citrate in this process, it has been hypothesized that enzymes involved in citrate metabolism may play a role in tumors (Santos and Schulze, 2012; Currie et al., 2013). Citrate is generated in the first step of the TCA cycle via a condensation reaction between acetyl-CoA and oxaloacetate catalyzed by the *CS* genes and is considered the rate limiting step in the Kreb's cycle. It has been shown in cervical carcinoma cells that knockdown of *CS* causes epithelial-mesenchymal transition resulting in increased proliferation, metastasis and high dependency on the Warburg effect like phenotype (Lin et al., 2012). However, in ovarian cancer cells *CS* knockdown led to decreased proliferation, invasion and drug resistance suggesting tumor specific metabolic adaptation upon *CS* perturbation (Chen et al., 2014). Moreover, expression of the citrate transporter *SLC25A1* is elevated in many tumor cells which correlates with poor patient survival. Genetic knockdown of *SLC25A1* in breast cancer cells *in vitro* causes reduced cell viability by disrupting mitochondrial homeostasis and causes reduced tumor development *in vivo* (Catalina-Rodriguez et al., 2012).

Lipids are among the most important constituent of cellular biomass and is actively produced in rapidly proliferating cells

A crucial central metabolite in both anabolic and catabolic processes of lipid metabolism is acetyl-CoA. It is generated within the cell by multiple metabolic pathways like amino acid degradation, glycolysis and β-oxidation of fatty acids. Acetyl-CoA serves as the primary precursor molecule which is then processed into the vast repertoire of lipids and fatty acids seen within a cell, furthermore it is also passed through the TCA cycle to generate cellular energy currency and reducing equivalents as ATP and NADPH/FADH respectively (Currie et al., 2013). Interestingly, in multiple tumor derived cells, it has been shown using isotope labeling experiments that a unique metabolic reprogramming event leads to the production of acetyl-CoA under oxygen limiting conditions (hypoxia). It is observed that α-ketoglutarate produced through glutaminolysis is converted into citrate via reductive carboxylation of glutamine us-

ing the cytosolic $NADP^+$ dependent *IDH1* enzyme. This reaction occurs in the opposite direction compared to the normal TCA cycle, hence is also called the "reversed TCA cycle" (Mullen et al., 2012; Metallo et al., 2012). This citrate pool is further converted into acetyl-CoA and oxaloacetate in an energy dependent reaction catalyzed by *ACLY*, this reaction acts as the crucial precursor step that connects carbohydrate metabolism to fatty acid biosynthesis (Fig. 6, 5). Moreover, this reductive

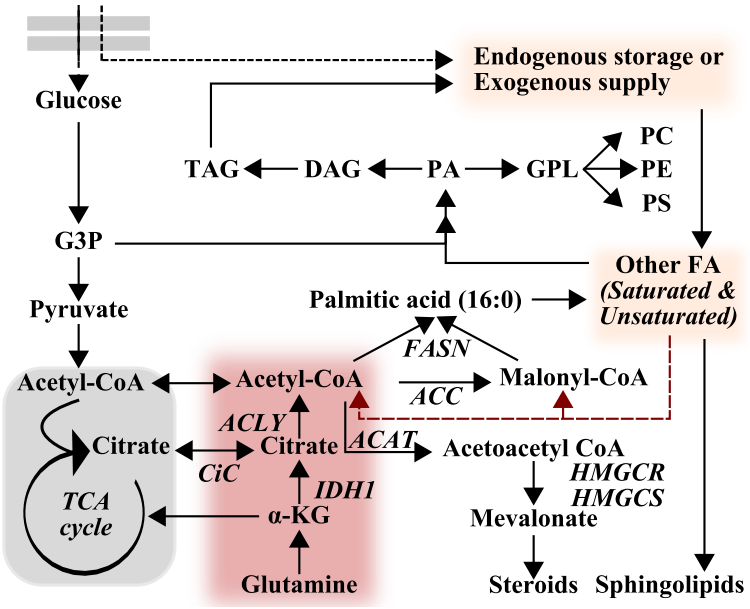


Figure 6: **Lipid metabolism in tumors.** A broad overview of various reactions involved in tumor's lipid metabolism is shown. Important enzymes altered in tumors are shown in italics. The red dotted line represents β-oxidation of fatty acids. The gray and pink shaded regions is used to differentiate the TCA cycle and the reverse TCA cycle (reductive carboxylation) respectively. G3P - Glyceraldehyde-3-Phosphate, α-KG - α-ketoglutarate, CiC - Citrate transporter, FA - Fatty acids, PA - Phosphatidic acid, DAG - Diacylglycerol, TAG - Triacylglycerol, GPL - Glycerophospholipids, PC - Phosphotidylcholine, PE - Phosphotidylethanolamine, PS - Phosphotidylserine.

carboxylation phenotype is also exhibited in renal carcinoma cells harboring a loss of function mutation in the *VHL* gene, where these cells become exclusively dependent on glutamine derived lipogenesis (Metallo et al., 2012). As expected, given the importance of *ACLY* in lipid metabolism, it has been shown in multiple tumor cells that genetic knockdown of *ACLY* causes cell cycle arrest and induction of apoptosis (Zaidi et al., 2012; Hanai et al., 2012). Furthermore, chemical inhibitors against *ACLY* in glycolytic tumors had similar effects (Hatzivassiliou et al., 2005).

Fatty acid synthase (FAS), encoded by the *FASN* gene, is the major multi-enzyme complex involved in lipid biosynthe-

sis. Mammalian FAS is a multifunctional protein consisting of 7 functional domains that iteratively carry out a set of 4 catalytic steps – condensation, reduction, dehydration and reduction. This process starts with condensation between acetyl-CoA and malonyl-CoA to form acetoacetyl-ACP (i.e. acetoacetyl conjugated with the acyl carrier protein) and ends with the final reduction step forming butyryl CoA-ACP. These set of 4 steps are repeated 7 times to generate palmitate (16:0), which then serves as the precursor for the vast range of fatty acids seen within the cells (Fig. 6)(Leibundgut et al., 2008). The initial condensation step, which is also the commitment step into fatty acid biosynthesis is catalyzed by *ACC* and is under tight signaling and allosteric regulation (Currie et al., 2013). In tumor cells, fatty acid biosynthesis is mainly carried out by *de novo* mechanisms. Unlike normal cells, where *FASN* exhibits tissue specific expression mainly in hepatocytes and adipocytes, it has been observed that *FASN* is highly expressed in multiple cancer types and its oncogenic role has been linked to poor survival prognosis (Menendez and Lupu, 2007). Furthermore, use of small inhibitors against *FASN* or its genetic knockdown leads to apoptosis and improves survival (Lupu and Menendez, 2006; Flavin et al., 2010), however direct pharmacological targeting of *FASN* is probably futile as it leads to severe side effects (Loftus et al., 2000).

2.3.4 *Regulation of tumor metabolism*

Metabolism is an essential and highly conserved cellular process that supplies the fuel required to sustain the survival and fitness of multicellular organisms. This process is tightly regulated by both intra- and extracellular environment through complex information signaling and metabolic homeostasis which maintains an optimal metabolic state for the organism. In rapidly proliferating cancer cells, various signaling programs interact and crosstalk eventually targeting downstream metabolic targets thus regulating their catalytic activity. It is increasingly being understood that metabolic remodeling is driven by increased oncogenic function and loss of tumor suppressive mechanisms (King et al., 2006; Jones and Thompson, 2009; Ward and Thompson, 2012). Often metabolites themselves regulate the flux through pathways via allosteric regulation which leads to flux control through both negative or positive feedback loops (Ros and Schulze, 2013; Gui et al., 2013). Also, metabolites can directly modify proteins and DNA via posttranslational modifications (like acetylation of histones or glycosyla-

tion of receptors) (Lin et al., 2014; Metallo and Vander Heiden, 2010) or epigenetic regulation (methylation of nucleotides) (Ulanovskaya et al., 2013). Therefore, rewired metabolic and signaling networks within the cell complement and coordinate with one another to feed the biosynthetic and bioenergetic demands of tumors thus maintaining its growth and proliferation.

Gain in oncogenic function

The major oncoproteins that regulate tumor metabolism are PI3K/Akt, Myc, Kras and *HIF1-α* interestingly all of their roles converge into regulating glucose/glutamine uptake and their metabolic transformations into subsequent products (Ward and Thompson, 2012; Cairns et al., 2011). The phosphoinositide-3-kinase/Akt (PI3K/Akt) signaling pathway plays a crucial role in glucose metabolism. The PI3K heterodimer, upon its activation through binding of extracellular growth signals to receptor tyrosine kinases (like *EGFR*), phosphorylates phosphatidylinositol-4,5-bisphosphate in the cell membrane to generate phosphatidylinositol-3,4,5-trisphosphate which acts as an anchor for Akt (a serine/threonine kinase). Akt mediates the downstream effects of activated PI3K by increasing glucose uptake and glycolytic flux (Elstrom et al., 2004; Engelman et al., 2006).

Glucose and glutamine metabolism are tightly regulated by oncogenes and tumor suppressor genes

Akt increases the translation and membrane localization for glucose transporters (like *GLUT1*). Furthermore, Akt enhances the activity of the glycolytic enzymes hexokinase (*HK1*) and phosphofructokinase (*PFK1*) (Deprez et al., 1997; Gottlob et al., 2001; Rathmell et al., 2003). Also, Akt increases the utilization of cytosolic acetyl-CoA for fatty acid biosynthesis by activating ATP-citrate lyase (*ACL*) (Berwick et al., 2002). Moreover, Akt activates the Sterol Regulatory Element Binding Protein transcription factors (SREBP) which in turn activate genes like *ACLY*, *ACC* and *FASN* that are crucial in cholesterol and fatty acid biosynthesis (Krycer et al., 2010). Akt can further regulate crucial transcriptional factors like FOXO, mTOR and HIF1, the overall interplay among these oncoproteins causes increased glycolysis and protein biosynthesis (Hemmings and Restuccia, 2012; Ward and Thompson, 2012). The oxygen sensing transcription factor *HIF1-α* is active under hypoxic conditions and is an important player in the Warburg effect. *HIF1-α* itself is activated by mTOR and this causes increased glycolysis by enhancing glucose uptake and activating glycolytic enzymes (Cairns et al., 2011; Ward and Thompson, 2012; Schulze and Harris, 2012). Furthermore, *HIF1-α* activates lactate dehydrogenase (*LDHA*) that leads to higher lactate production and also inhibits pyruvate dehydrogense (*PDH*) by activating pyruvate dehydro-

genase kinase (*PDK*) (Semenza, 2011). Oncoproteins like Myc and Kras mainly affect glutamine uptake by over expressing glutamine transporters and regulating glutaminolysis enzymes. For example, in pancreatic ductal adenocarcinoma it has been shown that Kras regulates the non-cannonical glutaminolysis pathway by suppressing the expression of *GLUD1* and activating *GOT1* (Son et al., 2013). Similarly *MYC* can activate the enzyme glutaminase *(GLS1)* involved in conversion of glutamine into glutamate while overexpressing high affinity glutamine transporters (like *SLC38A5* and *SLC1A5*) involved in glutamine uptake (Wise et al., 2008). Moreover, Myc and Kras are also involved in regulating glycolysis by activating among other glycolysis enzymes, *LDHA* and *GLUT1* respectively (Shim et al., 1997; Yun et al., 2009). Myc and Kras oncoproteins also play a prominent role in nucleotide biosynthesis, it has been shown that Kras drives excess glycolytic carbons into the non-oxidative arm of the pentose phosphate pathway to generate riboses required for nucleotide biogenesis (Ying et al., 2012) while Myc directly regulates the expression of genes (like *PRPS2, TS, IMPHD1/2, CAD*) involved in both purine and pyrimidine biosynthesis (Miltenberger et al., 1995; Mannava et al., 2008; Liu et al., 2008; Cunningham et al., 2014). Constitutive activation of these oncogenes due to mutations drastically transforms the metabolic landscape of tumor cells towards increased biomass production through glycolytic and glutaminolytic intermediates that are important precursors for various anabolic processes.

Loss of tumor suppressive function

Tumor suppressor genes are involved in maintaining the cellular homeostasis, a genomic disruption of such genes may lead to the cancer phenotype. In context of tumor metabolism, multiple tumor suppressor genes are disrupted to overcome the regulatory effects exerted by these genes as a result of which cell undergo rapid proliferation by tuning metabolism to increase the rate of biomass synthesis (Jones and Thompson, 2009). It has been shown that the master regulator p53 controls flux through glycolysis by activating *PTEN* which is a negative regulator of PI3K/Akt pathway as a result of which the rate of glycolysis is reduced (Stambolic et al., 2001). Furthermore, it can directly regulate glycolysis by inhibiting the expression of glucose transporters (*GLUT1/4*) (Schwartzenberg-Bar-Yoseph et al., 2004). Also, p53 controls glycolysis through TP53-induced glycolysis and apoptosis regulator (TIGAR) that specifically inhibits fructose-2,6 bisphosphate, a crucial intermediate in glycolysis (Bensaad et al., 2006). Interestingly, wild type p53 pro-

motes oxidative phosphorylation in normal condition by activating *SCO2* which is required for cytochrome c oxidase function in the electron transport chain (Matoba et al., 2006). Activated p53 also increases beta-oxidation by over expressing *CPT1*, a crucial transporter of fatty acids in the mitochondria (Zaugg et al., 2011). Another important tumor suppressor is *LKB1* which regulates metabolism by activating AMPK which is an important energy sensor of the cell (Shackelford and Shaw, 2009). It functions antagonistic to mTOR leading to *HIF1-α* inhibition, consequently, the rate of glycolysis and protein synthesis is reduced (Shackelford et al., 2009; Faubert et al., 2014). This effect on glycolysis control is further reinforced in nutrient and energy limiting conditions where AMPK acts as an activator of p53 (Vousden and Ryan, 2009). In some cases metabolic genes themselves have tumor suppressive function like the two consecutive metabolic genes in the TCA cycle, succinate dehydrogenase (SDH) and fumarate hydratase (FH). It has been suggested that mutations in these enzymes leads to accumulation of their catalytic products, succinate and fumarate that can stabilize *HIF1-α* under normoxic conditions through interactions with *HIF1-α* regulator prolyl hydrolylase (*PHD*) (King et al., 2006). In tumors, these metabolic checkpoints regulated by tumor suppressor genes are disrupted due to genomic aberrations in these genes as a result of which cancer cells go into overdrive, rapidly generating essential building block for their survival and proliferation.

Allosteric regulation and oncometabolites

Metabolites themselves can regulate the production and consumption of other metabolites in the cell through allosteric regulation by binding to a non-active site of metabolic enzymes and regulating their activity. In many instances, tumor metabolism is directly regulated by cellular metabolite levels suggesting its important role in non-canonical (non-signaling based) regulation of tumor metabolism. A well studied example is the allosteric regulation of a key glycolysis enzyme pyruvate kinase M2 isoform *(PKM2),* implicated in causing the Warburg effect. It has been shown that *PKM2* (a less active catalytic isoform) decreases the rate of glycolysis as a result of which crucial intermediates of this process are diverted into biosynthetic pathways leading to increase in biomass production. *PKM2* catalyzes the final step of glycolysis in its active tetrameric form where phosphoenolpyruvate is converted into pyruvate (Vander Heiden et al., 2010). It has been shown that fructose-1,6-bisphosphate (a product of glycolysis) and serine (a anaplerotic

product of glycolysis) acts as strong activators of *PKM2* function by stabilizing its active enzymatic structure (Dombrauckas et al., 2005; Chaneton et al., 2012; Gui et al., 2013)while L-alanine (which is generated from pyruvate, a *PKM2* catalyzed product) is shown to inhibit *PKM2* activity (Van Veelen et al., 1977). In nucleotide metabolism, maintenance of deoxynucleoside triphosphate (dNTP) pools is of critical importance to cancer survival. It has been shown that the dNTP triphosphohydrolase (*SAMHD1*) and ribonucleotide reductase (*RNR*), involved in the salvage and *de novo* nucleotide biosynthesis is tightly regulated allosterically by dNTP concentrations. Interestingly, as these enzymes have opposing functions, dNTP pools regulate their complementary functions through a feedback loop (Kohnken et al., 2015). Sometimes metabolites can directly regulate signaling proteins, for example, a low ATP:AMP ratio (energy deficit condition) is the activating signal for AMPK which regulates a plethora of signaling programs that control metabolism (Cairns et al., 2011). Similarly, the TCA cycle products succinate and fumarate can directly activate *HIF1-α* by inhibiting PHD (a negative regulator of *HIF1*) through competitive inhibition with α-ketoglutarate as described in the previous section (King et al., 2006). Such metabolites implicated in direct oncogenic roles are called "oncometabolites". A well studied example is the gain of function mutation in isocitrate dehydrogenase (*IDH1/2*) enzyme first observed in gliomas which leads to the increased production of D-2-hydroxyglutarate (2-HG). 2-HG competitively inhibits the binding of α-ketoglutarate with α-ketoglutarate dependent dioxygenases like PHD thus affecting *HIF1-α* dependent signaling (Cohen et al., 2013). Moreover, 2-HG causes widespread DNA methylation changes (hypermethylation phenotype - G-CIMP) by interacting with histone demethylases (KDM's) and DNA hydroxylases (TET) leading to vast changes in the expression of genes involved in cell differentiation and malignancy (Yang et al., 2012; Turcan et al., 2012).

Metabolites can act as non-canonical regulators of the metabolic process

2.4 GENOME ORGANIZATION AND ITS ROLE IN CANCER

Genes are broadly organized within the genome at linear and spatial levels. At both these levels of genome organization, it has been shown that proximal genes interact which results in phenotypic effects. Linear organization refers to the sequential arrangement of genes in each chromosome i.e. in one dimension and is also called the gene order, while the spatial organization is the position of the genes with respect to each other in three dimensions. Spatial arrangement of genes can be further

differentiated as intra-chromosomal or inter-chromosomal organization. Intra-chromosomal associations between genes occur due to genome organization emerging from chromatin dynamics events (within single chromosomes) like chromatin looping, while inter-chromosomal association among genes arises from the arrangement of freely diffusing chromatin structures within the boundaries of the cell nucleus (Hurst et al., 2004). In this section, the main focus will be on the linear gene organization and its influence on gene expression. Moreover, the regulation of these proximal gene clusters, its phenotypic effects and the role of genome evolution on such organizational designs will be explored. In addition, the role of gene positioning in the genome and its effect on the proliferative fitness of cancer cells will be described, supporting the emerging idea in cancer genomics that cancer is not necessarily the consequence of binary states – driver and neutral mutations, but rather a resultant product of additive gene dosage effects (Solimini et al., 2012; Davoli et al., 2013).

2.4.1 *Linear gene organization and expression*

Gene order is non-random, proximal genes are co-expressed and are enriched in functionally related genes

Before the complete sequencing of eukaryotic genomes was feasible, the gene order was considered to be random and independent of the gene expression patterns. However, certain exceptions to this perceived notion of random gene order were known, like the Hox and β-globin gene clusters that arose from gene duplication events. Now, with the availability of high resolution gene maps for large numbers of eukaryotic genomes, it is firmly established that gene order is actually non-random. Furthermore, it has been observed that the linear arrangement of genes in the genome affects gene expression independent of shared promoter sequences or transcription factors. Interestingly, in all major eukaryotic model systems like *S.cerevisiae, C.elegans, D.melanogaster* and human derived cells, linearly proximal gene clusters in the chromosome exhibit coordinated co-expression patterns and often belong to similar functional groups and pathways (Hurst et al., 2004). Chromatin dynamics seems to play a major role in co-expression of proximal genes. Chromatin can exist in euchromatic (transcriptionally active) or heterochromatic (transcriptionally inactive) states and it has been argued that transcriptionally active, open chromatin structures causes a ripple effect on neighboring genes leading to coordinated gene expression (Ebisuya et al., 2008). Furthermore, it has been seen that histone proteins bound at these regions have increased acetylation levels (Sproul et al., 2005; Batada

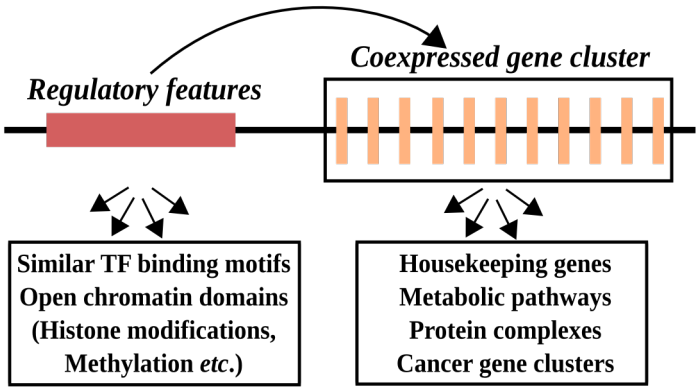


Figure 7: **Gene order and co-expression gene cluster.** Shared regulatory features control the coordinated expression of proximal gene clusters which also exhibit co-functionality.

et al., 2007). Co-expression of gene clusters is also attributed to sharing of common regulatory elements (Fig.7) (Davila Lopez et al., 2010; Trinklein et al., 2004; Wei et al., 2011; Semon and Duret, 2006). An interesting cause of gene expression is the "bystander effect" where a gene may be expressed solely because of its favorable position proximal to a highly expressing gene. For example, it has been shown that the gene *CD79b*, a B-cell specific expressing enzyme, is located between the human growth hormone cluster (*hGH*) and its regulatory element, the locus control region (*hGH* LCR). *hGH* is a pituitary specific expressing gene which influences the coordinated expression of *CD79b* in pituitary through *hGH* LCR, independent of the transcription factors required for *CD79b* expression in its native B-cells where surprisingly *hGH* LCR has no role in *CD79b* expression (Cajiao et al., 2004; Ho et al., 2006). Also, co-expressed gene clusters exhibit co-functionality i.e proximal genes often play similar functional roles. Housekeeping genes are associated with crucial cellular processes that are generic to multiple tissue types and often exhibit a similar expression pattern across tissues, such genes were observed to be positionally clustered in the genome and exhibited strong gene co-expression (Lercher et al., 2002). Similarly, it has been shown that 68% of all metabolic pathways annotated from Kyoto Encyclopedia of Gene and Genomes (KEGG) are clustered in the in human genome. Furthermore, adjacent gene pairs are significantly enriched in similar GO terms (Lee and Sonnhammer, 2003; Vogel et al., 2005; Al-Shahrour et al., 2010). Additionally, it has been shown that genes coding for co-functional proteins belonging to the same complex or interacting proteins in protein-protein interaction networks are localized on fewer chromosomes and

proximally positioned in the same chromosome than expected by chance (Thévenin et al., 2014). Multi-species studies have shown that the cause for such co-expressed proximal gene clusters is natural selection, suggesting that these observed features are functional due to evolutionary selection (Singer et al., 2005). Moreover, it has been shown that broadly expressing (similar expression across multiple tissue types) and co-expressed proximal gene clusters are under purifying natural selection based on observations that highlight a lack of chromosomal breakpoints in these gene clusters between mouse and human genomes (Semon and Duret, 2006; Purmann et al., 2007).

2.4.2 *Role of gene order in cancer*

A classical model for carcinogenesis is the Kundson two hit hypothesis in which two successive mutations are required for a recessive tumor suppressor gene like *Rb* to initiate cancer (Knudson, 1971). Similarly, for a dominant oncogene, a single mutation is sufficient to cause tumorigenesis. In many tumors a simple sequential steps of genetic aberrations causes tumorigenesis, for example in colorectal carcinomas successive alterations involving *APC, KARAS, SMAD2/4* and *TP53* leads to tumor formation (Fearon, 2011). These observations have driven the concept of "driver" and "passenger" mutations in cancer genomics, where mutations arising in a handful of cancer causing genes are sufficient to drive cancer and the large number of other observed mutations are simply neutral with no functional effect (Vogelstein et al., 2013). However, a recently proposed model of cancer genomics is that of "cumulative haploinsufficiency and triplo-sensitivity", in which cancer growth is rampantly affected by the cumulative gene dosage effects of multiple haplo-insufficiency (loss of an allele) or triplo-sensitivity genes (gain of an allele). It was shown that additive gene dosage effects from many passenger mutations (but not individually) can exert a strong effect on tumorigenesis (Solimini et al., 2012; Davoli et al., 2013). Furthermore, it has been shown that regions of hemizygous deletions in the cancer genomes are enriched for genes that negatively regulate (STOP genes) and depleted for genes that positively regulate (GO genes) cancer cell proliferation, such regions are called "cancer gene islands" that undergo hemizygous deletions and optimize the cancer fitness through cumulative haploinsufficiencies (Fig.8) (Solimini et al., 2012; Davoli et al., 2013).

Proximal genes in the genome are often co-altered by copy number changes and elicit a concerted functional phenotype.

It has been shown that in multiple epithelial malignancies like cancers of prostate, breast, colon and ovary, there exists

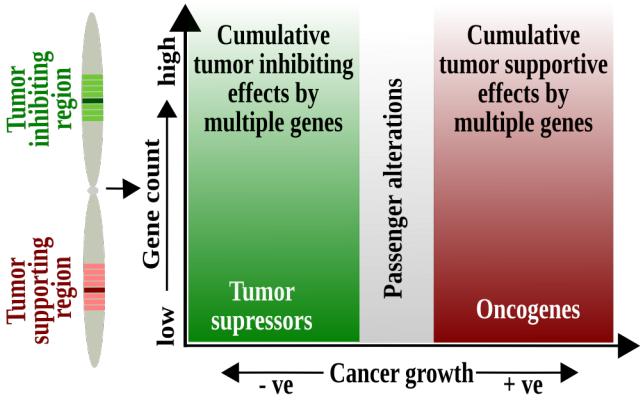


Figure 8: **Cumulative gene dosage model for cancer.** The concept of "cancer gene islands" and cumulative haploinsufficiency and triplosensitivity proposed by (Solimini et al., 2012; Davoli et al., 2013) is shown here. The main idea being that a large number of genes with altered gene dosage can cumulatively affect cancer potential as strongly as a single cancer driving oncogene or a tumor suppressing gene.

transcriptomeres i.e groups of cancer causing genes that tightly cluster together in the genome and are under co-transcriptional control (Glinsky et al., 2003). Furthermore, it has been observed in breast and ovarian carcinomas that tumor-related genes were enriched in having bidirectional promoters, many of which also shared common transcription factor binding sites (Yang et al., 2007). In melanoma cells, broad loss of heterozygosity due to deletions in chromosomes 6, 10 and 11 resulted in decreased expression of those regionally targeted genes and were involved in controlling tumor growth (Kwong and Chin, 2014). Similar synergistic and co-suppression effects were also observed in hepatocellular carcinoma for genes residing in chromosome 8. Down-regulation of these gene clusters (and not a single gene individually) correlated with poor survival (Xue et al., 2012). Studies using a functional screening approach to target genes amplified in hepatocellular carcinoma identified that *CCND1* (a well known driver gene) to be co-amplified with its proximal *FGF19* gene in chromosome 11 and established that the latter was an equally important driver that affected carcinogenicity and cancer growth. Additionally, in the same chromosome *BRIC2* and *YAP1* were identified as co-amplified and co-driving adjacent oncogenes (Zender et al., 2006; Sawey et al., 2011). In lung squamous cell carcinomas, it has been shown that knocking down the co-amplified proximal oncogene cluster in chromosome 3 consisting of *SENP2*, *DCUN1D1* and *DVL3* led to cancer growth inhibition. Furthermore, their expression levels enabled stratifying the patients based on adjuvant chemotherapy response (Wang et al., 2013). Also, in the

same cancer type and chromosome, an interesting mechanism of functional co-operation between co-amplified adjacent genes *PRKC1* and *SOX2* was observed. *PRKC1* phosphorylates *SOX1* and recruits it to the *HAAT* promoter region directly regulating Hedgehog signaling to maintain cancer stemness (Justilien et al., 2014). Clear cell renal carcinomas are typically characterized by deletion or loss of function mutation in the tumor suppressor gene *VHL*. In the context of tumor metabolism it has been shown that loss of heterozygousity in genes proximal to *VHL* affect metabolism. Furthermore, this co-deletion of multiple genes clustered together in chromosome 3 around *VHL* gene shapes the typical metabolic landscape observed in clear cell renal carcinomas which was found to be very distinct from all other cancers (Gatto et al., 2014). Similarly, in many tumors it has been shown that co-driving metabolic genes proximal to a well characterized cancer gene undergo copy number co-alterations leading to a functional effect on tumor metabolism. For instance in breast cancer *PNMT/ERBB2* and *PAK1/NDUFC2* are strongly co-amplified and are considered cancer co-driving genes (Bashashati et al., 2012). Likewise, the co-deletion of *CDKN2A* with *MTAP* has been long known example and is observed in multiple cancer types (Carson et al., 1988; Bertino et al., 2014).

3 METHODOLOGY

[1]We describe here the various methodological procedures used in the analysis of pan-cancer genomics data for identifying unique metabolic reprogramming events. Essentially, we integrated various data sources into a coherent analytical framework to study the dependency of metabolic alterations in tumors on copy number changes and gene proximities. The major task here was to filter out crucial co-driving metabolic genes from a vast set of copy number co-altered gene pairs. We assembled relevant data (cancer genomics and *a priori* biological knowledge) from various large datasets for both exploratory and targeted analysis. Furthermore, we gleaned novel insights from these integrated data sources by developing a statistical analysis pipeline, constructing a gene centric metabolic network and performing robust statistical tests to validate our predictions.

3.1 ASSEMBLY OF DIFFERENT DATASETS

3.1.1 *Assembling protein coding genes*

Our analysis involved measuring specific gene pair distances and identifying genes that led to measurable phenotypic effects, thus we focused on protein coding genes with defined chromosomal locations. A unique set of well annotated known protein coding genes with unambiguous (i.e having matching Entrez Gene IDs and HGNC symbols both in Biomart and HUGO databases) gene identifiers were downloaded from Ensembl Biomart and HUGO (GrCh38 release, downloaded on 12th Aug, 2014). To avoid major gender biases, only genes from autosomal chromosomes were selected for further analysis. Additionally, genes without a specific annotated chromosomal location were removed. This resulted in a total of 17,464 protein coding genes. This dataset was primarily used for three purposes – (1) to map gene annotations from all other datasets used in this study to maintain consistency (2) to compute all gene pair distances between genes from each chromosome and (3) obtain positional information for every cancer and metabolic genes in each chromosome. Furthermore, we also downloaded from Biomart, the isoform count for each of these protein-coding genes and ex-

1 This chapter has been taken from (Sharma et al., 2015)

tracted the rates of synonymous (dS) and non-synonymous (dN) substitutions for a subset of these genes that were orthologous between human and the mouse genome. Isoform count and nucleotide substitution rates were used later in our study to identify similarities between cancer and metabolic genes based on these genic properties.

In one dimension (i.e. along the length of a chromosome), the Euclidean distance D between any two gene pairs (i, j) within a chromosome having midpoints gm and its start and end positions g_{start}, g_{end} respectively, can be calculated as –

$$gm = \frac{g_{start} + g_{end}}{2} \tag{1}$$

$$D_{ij} = \sqrt{(gm_i - gm_j)^2} \tag{2}$$

using (1) and (2), the Euclidean distance matrix M_d for all gene pairs in a chromosome is given by –

$$M_d = \|gm_i - gm_j\|_2^2 \tag{3}$$

Gene pair distance information was used to measure differences between cancer-metabolic and cancer-nonmetabolic gene pairs, to assess concordant effects on copy number co-alterations and to threshold co-altered gene pairs separated by less than 1Mb distance. The 1Mb cutoff was selected because it was recently shown in a pan-cancer somatic copy number alteration (SCNA) study that the average size of focal SCNAs away from the DNA telomeres in both amplification and deletion events is less than 1Mb (Zack et al., 2013).

3.1.2 *Somatic copy number variations and expression data*

Copy number variations and microarray or RNAseq based transcriptome expression data was downloaded from the Firehose portal hosted at the Broad GDAC genomic data repository of the Broad Institute (`http://gdac.broadinstitute.org/`, release 2014-07-15) for 19 different cancer types (Fig. 9 and Table 2). These datasets had no restriction on usage according to the The Cancer Genome Atlas (TCGA) publication guidelines of 3rd Dec, 2014 (see `http://cancergenome.nih.gov/publications/publicationguidelines`). GDAC is a comprehensive repository of all cancer genomic datasets originating from the TCGA project followed by standardized processing pipelines. In this study we have used preprocessed copy number data generated using the Genomic Identification of Significant Targets in Cancer 2 (GISTIC2) algorithm, GISTIC2 identifies chromosomal re-

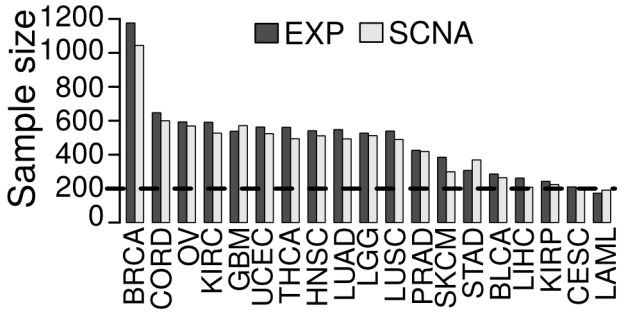


Figure 9: **Sample sizes for the TCGA data.** Sample sizes for 19 different cancer types used in our analysis, for all cancers there were more than 200 samples (dashed line) except for acute myeloid leukemia (LAML) (samples with LAML expression data (EXP) N=173 and copy number data (SCNA) N=191)

gions with significantly altered copy numbers by incorporating improved information on background alteration rates (Mermel et al., 2011). It does so by estimating the background alteration rates separately for focal or arm level SCNA events. This is crucial because it has been shown that SCNA frequency varies inversely with length (Zack et al., 2013), thus it is important to consider this inherent bias while predicting significantly changed gene copy numbers. The GISTIC2 derived data consisted of discretized values of -2, -1, 0, +1, +2 representing high level deletions, low level deletions, no copy number change, low level amplifications and high level amplifications, respectively. In all our analysis, we have used only high level deletions (-2) or amplifications (+2) as an indication of copy number alterations for a gene. For expression data, the platform (Agilent, Affymetrix, Illumina gene arrays or RNAseq) with the highest sample coverage was chosen . Pre-processed microarray data were used, that were RMA normalized and $\log_2$transformed while for RNAseq data, gene wise averaged $\log_2$RSEM values that were used, except for stomach adenocarcinoma (STAD). For STAD, RSEM values were unavailable therefore, the gene averaged and normalized $\log_2$RPKM values were used. In total, data from 8,515 and 9,116 cancer genomes were used for copy number and gene expression analyses respectively.

3.1.3 *Assembling a set of cancer causing genes*

The list of cancer causing genes was assembled using the following four different sources: (1) Cancer gene census (CGS) from the Catalogue of Somatic Mutations in Cancer (COSMIC). The list was downloaded from `http://cancer.sanger.ac.uk/cancergenome/projects/cosmic` on 13th Jan, 2014 (N=391). The

Table 2: **Abbreviations and sample sizes for the TCGA data†.**

Cancer	Description	Expression data	SCNA data
BRCA	Breast cancer	1176	1044
CORD	Colon & rectal adenocarcinoma	647	600
OV	Ovarian carcinoma	593	569
KIRC	Clear cell kidney carcinoma	591	527
GBM	Glioblastoma multiforme	538	571
UCEC	Uterine corpus endometrial carcinoma	563	524
THCA	*Papillary thyroid carcinoma*	*561*	*494*
HNSC	Head and neck squamous cell carcinoma	541	511
LUAD	Lung adenocarcinoma	548	493
LGG	Lower Grade Glioma	527	512
LUSC	Lung squamous cell carcinoma	539	490
PRAD	Prostate adenocarcinoma	426	419
SKCM	Cutaneous melanoma	385	299
STAD	Stomach adenocarcinoma	307	369
BLCA	Urothelial bladder cancer	286	264
LIHC	Liver hepatocellular carcinoma	262	208
KIRP	*Papillary kidney carcinoma*	*243*	*224*
CESC	Cervical cancer	210	206
LAML	*Acute Myeloid Leukemia*	*173*	*191*
	Total	9116	8515

†Complete names for abbreviations (standard TCGA cancer identifiers) used for the cancer types in this study and its respective sample sizes for expression and copy number data (also see Fig. 9). The standard identifier for colon and rectal adenocarcinoma is CORDREAD, which for simplicity is abbreviated as CORD here. For GBM and OV microarray expression data was available for more samples than RNAseq data, hence in our study we selected Affymetrix based microarray data for these two cancer types. For the remaining cancers RNAseq data was used. The cancers shown in italics (LAML, KIRP and THCA) did not have any gene pairs that passed our selection criterion during the statistical prioritization step and were not used for the downstream functional prioritization steps.

cancer gene census is a manually curated database which aims to catalog all genes implicated in carcinogenesis from experimentally verified and literature sources. (2) The Tumor suppressor and Oncogenes Explorer (TUSCON) (N=244) provided a list of novel tumor suppressor genes and oncogenes. We selected only those cancer genes with a false discovery rate (FDR) <0.05 from the list provided by Davoli and coworkers. Within this computational method, more than 8,000 tumor and normal genome sequences were analyzed to identify 22 parameters that could strongly predict known cancer genes, using the best predictors they identified novel cancer causing genes (Davoli et al., 2013). (3) A list of known tumor suppressor genes were taken from the TSGene database, downloaded on 13th Jan, 2014 (N=592). This database provided a list of tumor suppressor genes with literature evidence (Zhao et al., 2013). (4) We further used a list of cancer genes published in (Vogelstein et al., 2013) which provides a set of well known, high confidence cancer drivers (N=106). The union of these sources (1–4) was taken resulting in 1,065 cancer causing genes (Fig. 10A).

3.1.4 *Assembling a set of essential genes*

Our analysis required means for functionally prioritizing metabolic genes, identifying those that could have a major functional impact. Essential genes are indispensable for cellular survival and often involved in central cellular processes. We assembled a list of essential genes from four different sources, i.e. from (1) the Database for Essential Genes (DEG). This database lists essential genes from different organisms that have been experimentally verified. Only human specific essential genes were selected (N=101) (Luo et al., 2014); (2) the Online Gene Essentiality Database (OGEE), it provides essential genes identified from the literature for several organisms. Here again we selected genes being essential in human (N=1,380) (Chen et al., 2012); (3) genes from (Georgi et al., 2013) in which an evolutionary conservation approach was followed to map essential genes between mouse and human orthologues (N=2,331), and (4) the list of essential cancer genes identified across a panel of cancer cell lines, we selected only the list of "core essential genes" (N=749) (Hart et al., 2014). A union of genes from these sources resulted in a total of 3,839 unique essential genes which was used for functional association analysis with our list of predicted metabolic cancer genes (Fig. 10B).

3.1.5 *Human metabolic genes*

We extracted the list of metabolic genes (N=1,570) from the Kyoto Encyclopedia of Gene and Genomes (KEGG) restricting to human metabolic pathways (N=92 pathways). The KEGG database was accessed using the Bioconductor package KEGGREST (version 1.2.2). iPATH2 is a web based tool for visualizing metabolic pathways from KEGG, furthermore external data can be used to annotate these metabolic maps. We used iPATH2 for mapping copy number altered metabolic genes (identified using our analysis pipeline) into the global human metabolic network to visualize metabolic subsystems/modules targeted by these genes (Yamada et al., 2011). It has been shown that metabolic genes coding for enzymes catalyzing biochemical reactions at crucial nodes in a metabolic pathway majorly control the flux through the whole pathway. These nodes are usually at the beginning or end of a pathway and are often under tight transcriptional control. A list of such bottleneck genes for the human metabolic network (N=277) was provided by the authors of this study which was used in our analysis for functional validation studies of our predicted metabolic cancer genes (Wessely et al., 2011). The overlap of metabolic genes with other functionally defined gene lists used in this study is shown in Fig. 10C.

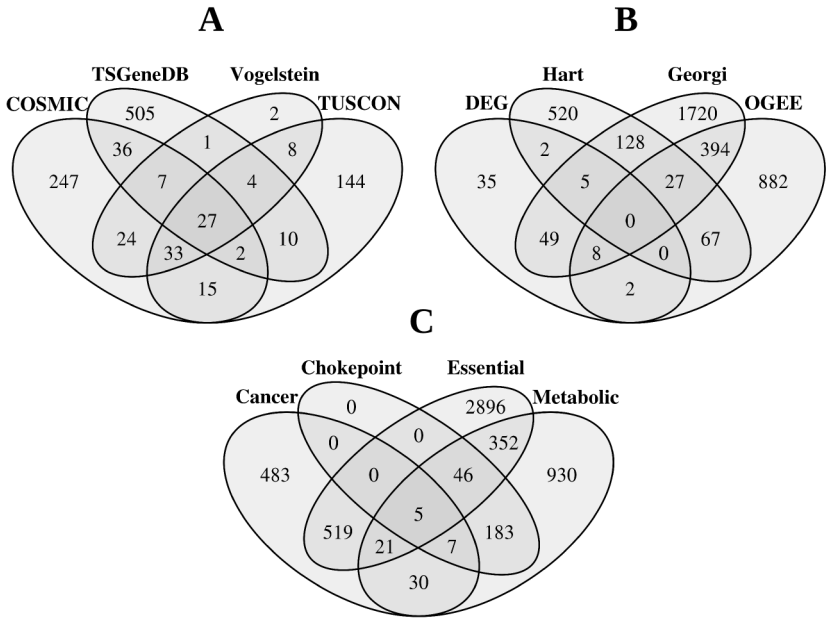


Figure 10: **Venn diagrams representing the overlap between various *a priori* defined gene lists.** Datasets used for annotating (A) cancer causing genes (B) and essential genes. (C) Overlap between various gene lists used for identifying metabolic functionality.

3.2 THE METABOLIC GENE CENTRIC NETWORK

We constructed a metabolic gene centric network to analyze how often a copy number co-altered metabolic gene also has the expression of its neighboring network genes changed between the groups of cancer samples defined by the copy number status of the former. A metabolic network G can be represented as a bipartite graph consisting of M metabolites as nodes and R reactions as edges –

$$G = (M, R) \tag{4}$$

where,

$$M = \{m_1, m_2, \cdots, m_M\} \tag{5}$$

and,

$$R = \{r_1, r_2, \cdots, r_R\} \tag{6}$$

In the metabolic network, a pair of reactions, for example $r_{1,}r_2$ will be connected if the substrate of one reaction served as the product of the other and/or shared common metabolites. We used such a metabolic network construction from KEGG and converted it into a reaction centric network representation by interchanging the nodes and edges of the network. We further mapped these reactions to their corresponding genes yielding a gene centric metabolic network (Fig. 11).

We used the KEGGREST package from Bioconductor, to extract data from the KEGG database pertaining to specific pathways of human metabolism for constructing this network. Initially for each metabolite, its connectivity (number of genes directly connected to a metabolite) to genes was measured and those with connectivity $\geq$45 were discarded, (except L-glutamate). These discarded hub metabolites (N=24) were mainly pool metabolites (like H_2O, H^+, O^2 etc.) and common co-factors (like *ATP, NADPH, CoA* etc). Using the remaining set of 1,851 metabolites, a gene centric metabolic network with 1,381 genes was constructed having at least 2 connections (avoiding dangling ends). The connectivity of this network ranged between a minimum of 2 and a maximum of 143 with median connectivity of 29. The connectivity value for each metabolic gene was later used in our functional analysis to identify if our predicted metabolic cancer genes had higher network connectivity than others. The statistical and graph properties of this metabolic network which represents a scale free network is elaborated in Appendix Table. 5.

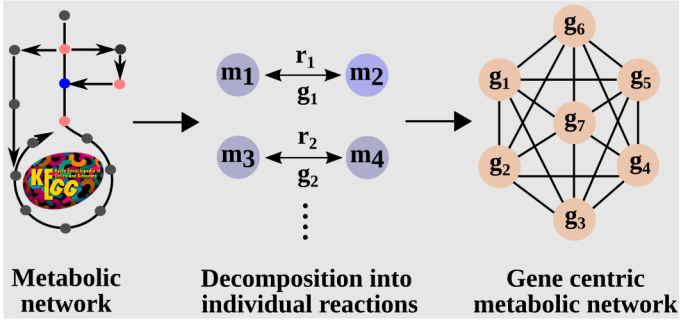


Figure 11: **Gene centric metabolic network construction.** A schematic representation of the metabolic network construction is shown which involves (i) extraction of data for human metabolic pathways from KEGG, (ii) decomposition of pathway information into their individual reactions and (iii) interchanging nodes into edges and mapping reactions to genes, such that two genes are connected if they share common metabolites.

3.3 THE ANALYSIS PIPELINE

The *iMetCG* analysis pipeline employed a two-fold prioritization process to identify functionally relevant co-altered cancer-metabolic gene pairs. The statistical prioritization step identified significantly co-altered gene pairs across cancer types. This was followed by a functional prioritization step, which filtered for gene pairs by comparing corresponding gene expression, evaluating expression perturbation in the metabolic gene network of neighboring genes and identifying metabolic genes that were present in *a priori* defined functionally important gene sets comprising of essential and bottleneck genes. A schematic representation of the *iMetCG* analysis pipeline is given in Fig. 12.

Statistical Prioritization

To identify cancer-metabolic gene pairs with significantly higher co-alteration events than expected by chance, we employed the method of pointwise mutual information (Bouma, 2009). For every investigated gene pair g_i, g_j in a cancer type with N samples, alteration events were coded as binary values. These binary values represented presence or absence of a high level SCNA. The statistical measure for the strength of co-alteration of g_i, g_j was calculated using a modified metric of pointwise mutual information *pMI* henceforward called the robust normalized *pMI* (*RnpMI*). *pMI* measures ratio of the probability of co-occurrence to the probability of occurrence for individual terms under the assumption of independence between the terms. Normalizing *pMI* yields *npMI* bounded to a fixed range of (-1, 1) making comparisons of information content among

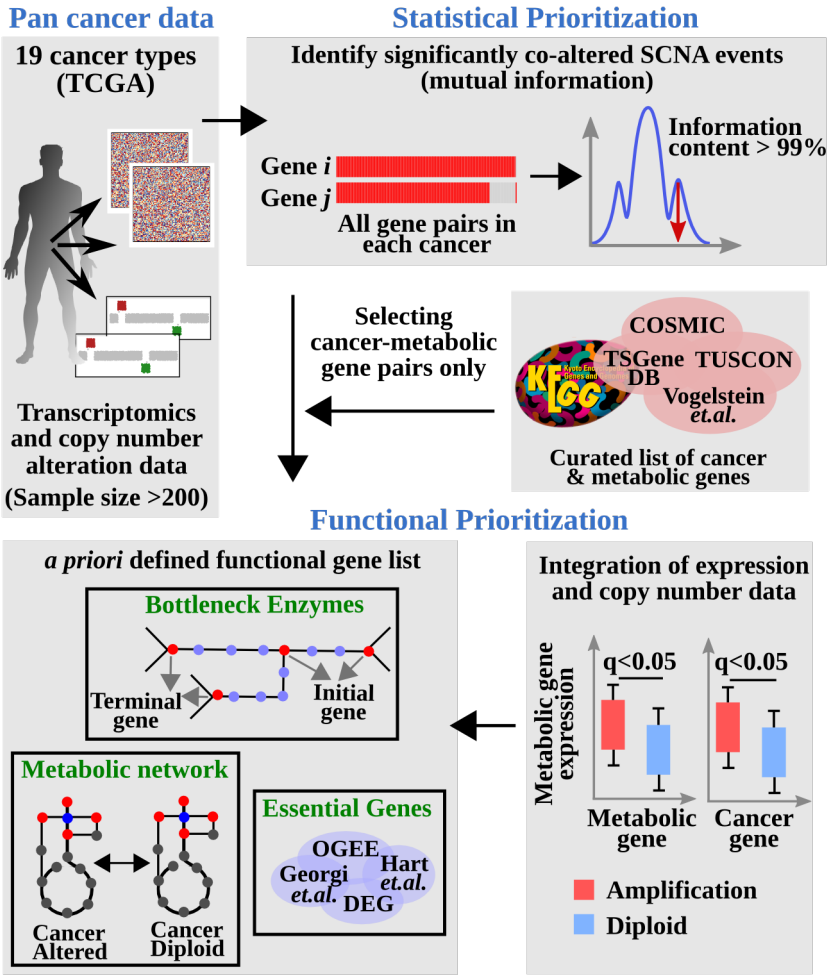


Figure 12: **The *iMetCG* analysis pipeline**. Schematic representation of the *iMetCG* pipeline. There are four major steps: (i) collection of transcriptomics and somatic copy number alteration data for 19 different cancer types from TCGA, (ii) statistical prioritization of all gene pairs in each cancer type, (iii) selection of cancer-metabolic (CG-MG) gene pairs and (iv) functional prioritization for the identification of putative metabolic cancer genes, which includes integration of expression and copy number data and further prioritization of hits using *a priori* defined biologically relevant functional gene lists.

co-altered gene pairs across cancer types easier to interpret. *RnpMI* is the normalized pointwise mutual information (*npMI*) value corrected for the low frequency problem (for details, see (Bordag, 2008)). In brief, the low frequency problem is that co-occurrences at low frequencies (being less likely) lead to the same *npMI* value as high frequencies (being more likely). To account for this, we used a method suggested by Washtell and coworkers and multiplied *npMI* by a frequency correction factor that takes the absolute frequency magnitude into consideration (Washtell and Markert, 2009), thus giving us a robust score for calculating significant co-occurrences, i.e.

$$RnpMI = \frac{pMI}{NF} \times CF = npMI \times CF \tag{7}$$

NF and *CF* denotes the normalizing factor and the low frequency correction factor respectively. Equation (7) corresponds to –

$$pMI = log\left(\frac{p\left(g_i = 1, g_j = 1\right)}{\left(p(g_i = 1) \times p\left(g_j = 1\right)\right.}\right) \tag{8}$$

$$NF = -log\left(p\left(g_i = 1, g_j = 1\right)\right) \tag{9}$$

$$CF = \sqrt{min\left(\left(p(g_i = 1) \times p\left(g_j = 1\right)\right)\right.} \tag{10}$$

Assembling all of which, gives us the full expression as –

$$RnpMI = \frac{log\left(\frac{p\left(g_i = 1, g_j = 1\right)}{\left(p(g_i = 1) \times p\left(g_j = 1\right)\right.}\right)}{-log\left(p\left(g_i = 1, g_j = 1\right)\right)} \times \sqrt{min\left(\left(p(g_i = 1) \times p\left(g_j = 1\right)\right)\right.} \tag{11}$$

with this, strongly co-altered cancer gene-metabolic gene pairs, with *RnpMI* scores higher than the 99% quantile from each cancer type were selected. Additional selection criteria were imposed, where we further filtered for gene pairs separated by less than 5Mb chromosomal distance and had at least 5% co-alteration frequency.

Functional prioritization

Functional prioritization was performed by considering the expression of the selected pairs of amplified cancer and metabolic genes. Firstly, the differential expression of the metabolic gene in its altered and diploid samples was calculated proceeded by calculating the differential expression of the metabolic gene in the sample sets with copy number altered and diploid cancer genes. Only pairs with significant differential expression in both case were selected ($p \leq 0.05$, corrected for multiple testing using the method given in (Benjamini and Hochberg, 1995)). To further filter and prioritize cancer-metabolic gene pairs, metabolic genes of functional importance were identified using our lists of essential genes and metabolic bottleneck genes. Furthermore, cancer-metabolic gene pairs were selected in which the nearest neighbors of the metabolic gene (in the metabolic

network) were enriched ($p \leq 0.05$) for differential expression between altered and diploid samples of the cancer gene.

3.4 STATISTICAL ANALYSIS

All p-values were calculated using a one-sided Wilcoxon rank sum test unless otherwise specified. Two sided Welch t-tests were used for differential gene expression analysis in functional prioritization steps (co-expression and network analysis). All enrichment tests were performed using one-sided Fisher's exact tests. Hierarchical clustering was performed using binary distances and the Ward's linkage method using the *dist* and *hclust* base functions of R. All data analysis, processing and statistics were carried out within the CRAN R statistical programming environment (`www.r-project.org`, version 3.0.0).

3.4.1 *Differential gene analysis*

We measured if gene expression of a metabolic gene was significantly different in its diploid (g_{dip}) and altered (g_{alt}) (i.e either amplified or deleted) copy number states. Additionally, difference in gene expression of the same metabolic gene was also measured between patients groups divided on copy number status of the proximal cancer gene. The co-altered metabolic genes which were differentially expressed in both cases above were selected. These were used to further prioritize those metabolic gene whose neighboring metabolic network genes also exhibited differential gene expression between patient groups stratified on altered (g_{alt}) or diploid (g_{dip}) copy number states for the proximal cancer gene. Differential gene expression analyses was performed using the Welch t-statistics (t). The Welch t-test unlike the commonly used Student t-test is suitable and more reliable when sample sizes and sample variations are unequal. However, the assumption of normal sample distribution is same for both tests. The Welch t-statistics is calculated as –

$$t = \frac{\overline{g}_{alt} - \overline{g}_{dip}}{\sqrt{\frac{v_{alt}^2}{n_{alt}} + \frac{v_{dip}^2}{n_{dip}}}} \tag{12}$$

where, v^2 and n represents the sample variance and sample size for the respective groups defined by their copy number status.

3.4.2 *Over-representation analysis*

Standard enrichment tests were performed to calculate the statistical significance for association between different variables. The patient samples were divided into two groups based on the copy number status of the cancer gene co-altered with a metabolic gene in the respective cancer type (see step (2 in 3.4.1) and the subsection functional prioritization above) and differential gene analysis was performed between these two groups. Let us suppose a to be the number of differentially expressed metabolic network neighbors, b to be the number of non-differentially expressed network neighbors, c to be the number of differentially expressed metabolic genes that are not network neighbors and d to be the number of non-differentially expressed metabolic genes that are also not network neighbors. Using these counts we performed Fisher's exact test to measure if metabolic network neighboring genes of the co-altered metabolic gene was significantly associated to the copy number status of the altered cancer gene partner.

Also, we measured the significance of known metabolic cancer genes (MCG) to be enriched in our list of predicted metabolic cancer gene (pMCG), similarly, let us suppose a to be the number of MCG in our list of pMCG, b to be the number of MCG not in our list of pMCG, c to be the number non-MCG metabolic genes in our list of pMCG and d to be the remaining metabolic genes that are neither MCG or pMCG.

Furthermore, we analyzed the association of pMCG to tumor survival, here we computed the survival difference for all metabolic genes (altered in at least 5% of the samples) in each cancer type between it copy number altered and non-altered groups using log-rank test and tested if the proportion of pMCG associated with metabolic genes had a significant survival difference while combining all cancer types. Again, like in previous cases, let us suppose a and b to be the number of pMCG and other metabolic genes (non pMCG) with significant survival difference while c and d are the number of pMCG and other metabolic genes (non pMCG) with no significant effect on survival. Using the counts a, b, c and d and the sum of all counts n, from each of the above 3 cases we constructed the confusion matrix M.

$$M = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \tag{13}$$

Using M, a one-sided Fisher's exact test was performed to calculate the probability p for the variables involved in each case to be significantly associated.

$$p = \frac{\binom{a+b}{a}\binom{c+d}{c}}{\binom{n}{a+c}} \tag{14}$$

3.4.3 *Clustering analysis*

Unsupervised agglomerative clustering analysis was performed using the Ward's linkage method (Ward Jr, 1963) to identify co-altered metabolic genes that were commonly targeted in multiple cancer types. We wanted to identify if co-altered metabolic gene signatures could be used to group tumors into distinct phenotypic sub-groups. Any clustering method requires a dissimilarity matrix for the attributes, therefore we constructed a binary matrix, $M_{g \times n}$ with g genes and n cancer types such that $M_{ij} = \{1, 0\}$, representing the presence/absence of a co-altered metabolic gene in that cancer type. For this matrix M, we then computed the pairwise binary distance matrix D between each pair of attributes (cancer samples) as –

$$D_{ij} = 1 - \frac{a}{a + b + c} \tag{15}$$

where a is the number of co-altered metabolic genes common between a pair of tumor types, b and c is the number of co-altered metabolic genes in one tumor but not the other and *vice verse* respectively.

4

RESULTS & DISCUSSION

[1]We compared distances between human chromosomal cancer-metabolic and cancer-nonmetabolic gene pairs, and statistically assessed co-alteration frequencies for proximal cancer-metabolic and cancer-nonmetabolic pairs. The *iMetCG* analysis pipeline was applied to each cancer type to identify co-altered cancer-metabolic gene pairs exerting a functional impact on cancer metabolism. We then prioritized cancer-metabolic gene pairs with high co-alteration and co-expression. These pairs were then functionally prioritized by selecting metabolic genes known to code for essential or bottleneck enzymes in a metabolic pathway. In parallel, we prioritized pairs for which expression perturbations extended to network neighbors of the metabolic genes in expression profiling data from that cancer entity. We investigated network connectivity, cancer survival and overlap with known cancer metabolic genes for the prioritized metabolic cancer genes. To evaluate the propensity to which these prioritized metabolic cancer genes are relevant for cancer, we further compared their selection pressure, evolutionary rate and isoform diversity to known cancer genes.

4.1 CANCER-METABOLIC GENE PAIRS ARE CLOSER TOGETHER IN THE GENOME THAN CANCER-NONMETABOLIC GENE PAIRS

The Euclidean distances between the protein-coding gene midpoints were calculated for each cancer-metabolic gene pair harbored together on a single chromosome. Using these gene pair distances calculated from all chromosomes (frequency counts for gene pairs across all chromosomes in Fig.13 and associated distance distribution statistics in Appendix Table 4), proximal pairs was defined as less than 1Mb apart. The 1Mb cutoff was selected because it was recently shown in a pan-cancer somatic copy number alteration (SCNA) study that the average size of focal SCNAs away from the DNA telomeres in both amplification and deletion events is less than 1Mb (Zack et al., 2013). Furthermore, the distribution frequency for gene pair distances measured across chromosomes revealed that there exits almost ~10 times more cancer-nonmetabolic than cancer-metabolic gene pairs. The distances between cancer-metabolic

1 This chapter has been taken from (Sharma et al., 2015)

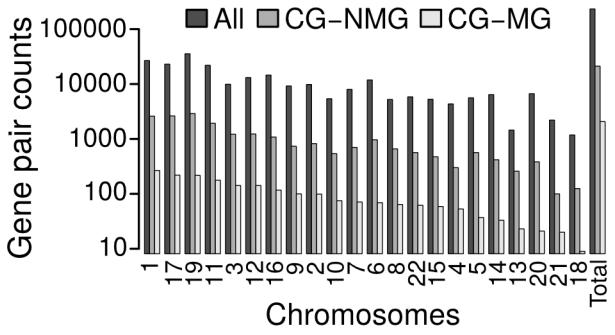


Figure 13: **Distribution frequency for gene pairs counts across chromosomes.** Gene pair counts in each chromosome for all gene pairs (ALL), cancer-nonmetabolic (CG-NMG) gene pairs and cancer-metabolic (CG-MG) gene pairs The y-axis has been drawn logarithmically and bars are sorted based on CG-MG gene pairs counts.

gene pairs were significantly smaller (p=0.023, median distance difference=33Kb) than those between cancer-nonmetabolic gene pairs (Fig. 14A). To assess whether the shorter separation between cancer-metabolic gene pairs that we observed did not occur by random chance, we shuffled the gene labels 10,000 times keeping the chromosomal gene density and gene occupancy constant. We re-computed the distance difference between cancer-metabolic and cancer-nonmetabolic gene pairs using the 1Mb cut-off in these randomized genomes, and found strikingly significant differences between our observed values and values calculated from the random genomes. None of these random differences were lower than our observed value of 33Kb (p=1e-04, Fig. 14B). Furthermore, we observed that 63% of all metabolic genes reside within 1Mb from all cancer genes. Within the same range, 83% of all metabolic pathways have at least 50% of its genes present proximal to a cancer gene. We observed a high coverage for several cancer-relevant metabolic pathways, such as oxidative phosphorylation (OXPHOS), purine and pyrimidine metabolism, glycolysis and synthesis of glycero-phospholipids and amino acids (Fig. 15B). We show the relative frequency of cancer and metabolic genes as fractions of all protein-coding genes for each chromosome as well as the distribution of their genomic locations in (Fig. 15A). From these data we conclude that on average, metabolic genes are distinctively closer to cancer genes than non-metabolic genes. The consequences of this observed linear contiguity on genomic co-aberration events is elucidated in the following sections.

Most cancer genes have metabolic genes in close proximity in the genome

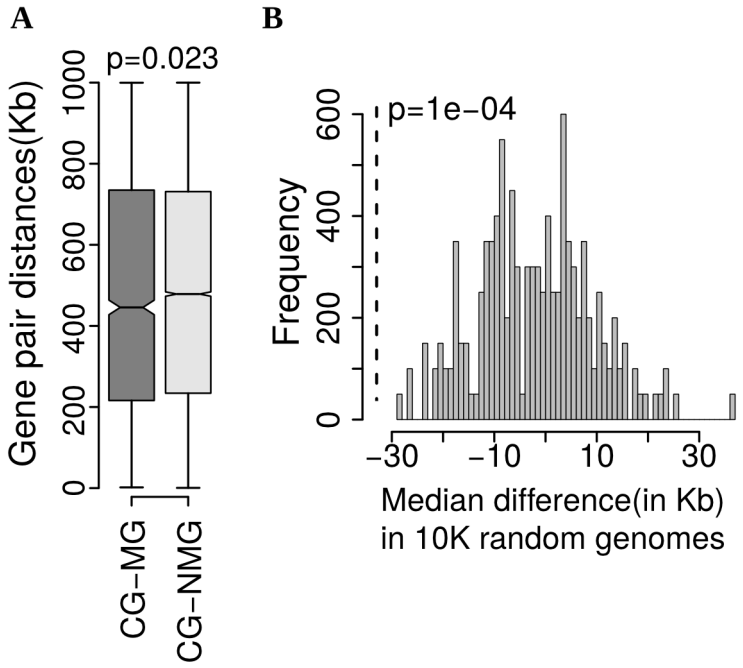


Figure 14: **Comparison of cancer-metabolic and cancer-nonmetabolic gene pair distances. (A)** Comparison of intra-chromosomal distances between all cancer-metabolic (CG-MG) and cancer-nonmetabolic (CG-NMG) gene pairs separated by <1Mb distance. **(B)** Permutation analysis using 10,000 random genomes (10K), CG-MG gene pair distances in the human genome were significantly shorter than in random genomes. The histogram shows the distribution for median distance differences between CG-MG and CG-NMG pairs in the random genomes, the black dashed line represents median distance difference in the real genome.

4.2 CANCER-METABOLIC GENE PAIRS HAVE HIGHER COPY NUMBER CO-ALTERATIONS

We investigated whether cancer-metabolic gene proximity also results in higher occurrences of cancer-metabolic gene co-alteration. We compared the co-occurrence of copy number alterations for cancer-metabolic and cancer-nonmetabolic gene pairs that were altered in at least one cancer sample across all 19 different cancer types studied. We selected a 1Mb cutoff and a minimum of 5% absolute co-occurrence frequency to observe SCNA effects within relevant gene pair distances. Co-alteration was measured by the information content of all selected cancer-metabolic and cancer-nonmetabolic gene pairs using normalized pointwise mutual information (*npMI*). A high *npMI* value corresponds to a high information content of co-occurring alterations, thus a highly non-random event. We observed significantly higher co-occurrences for cancer-metabolic gene pairs (p=1e-09, Fig. 16A). This is in accordance to the smaller cancer-

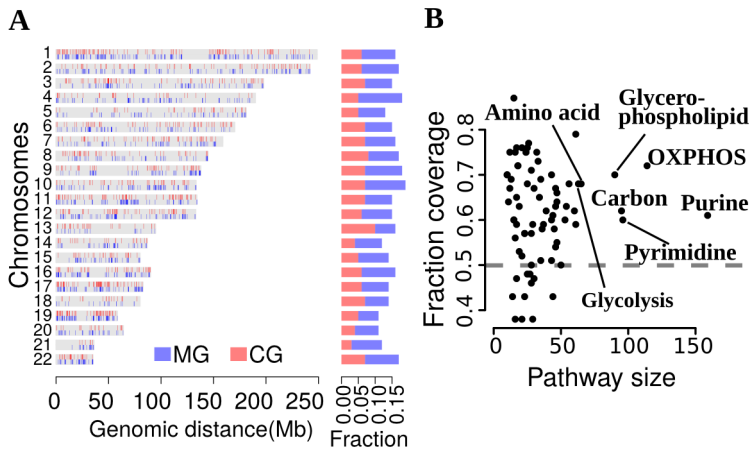


Figure 15: **The distribution of cancer and metabolic genes in the genome.** (A) The midpoints of all cancer genes (CG) and metabolic genes (MG) residing in its respective chromosomes are visualized. The fraction of protein coding CG and MG in each chromosome is shown on the right panel. (B) The coverage of MG (in context of metabolic pathways) within 1Mb distance of a CG is given. Pathway size is the number of genes in that metabolic pathway and coverage is the fraction of those genes within 1Mb distance to a cancer gene. Representative examples of pathways that have concurrently larger sizes and coverages are shown (amino acid biosynthesis, glycero-phospholipid metabolism, oxidative phosphorylation (OXPHOS), carbon metabolism, glycolysis, purine and pyrimidine metabolism), the gray dashed line represents a coverage of 50%.

Proximity of cancer-metabolic genes in the genome leads to higher rates of their copy number coalteration across cancer types

metabolic gene distances that we observed in the previous section. To further confirm our findings, we performed a binned distance analysis using the assumption that proximal genes in a SCNA-susceptible locus have higher chances of co-alteration. We binned the gene pair distances and compared cancer-metabolic and cancer-nonmetabolic gene pairs of similar distances (within 1Mb, using a sliding window of 250 Kb for 0-250Kb, 250-500kb, 500-750Kb and 750-1000Kb bins).We observed significant differences ($p \leq 0.05$), with mean cancer-metabolic gene pair *npMI* values higher than cancer-nonmetabolic gene pairs in all bins except the first with <250Kb separation (Fig. 16B). Our results support the fact that the proximity of cancer-metabolic gene pairs is exploited by cancer cells to alter metabolic gene copy numbers. We propose that these might not just be neutral bystander effects, but functional for at least a subset of these co-alterations where the affected metabolic gene produces a phenotype beneficial for cancer. These often co-occurring alterations targeting crucial cancer-relevant pathways could fulfill vital bio-energetic requirements for tumor cells, which we investigated in the next sections.

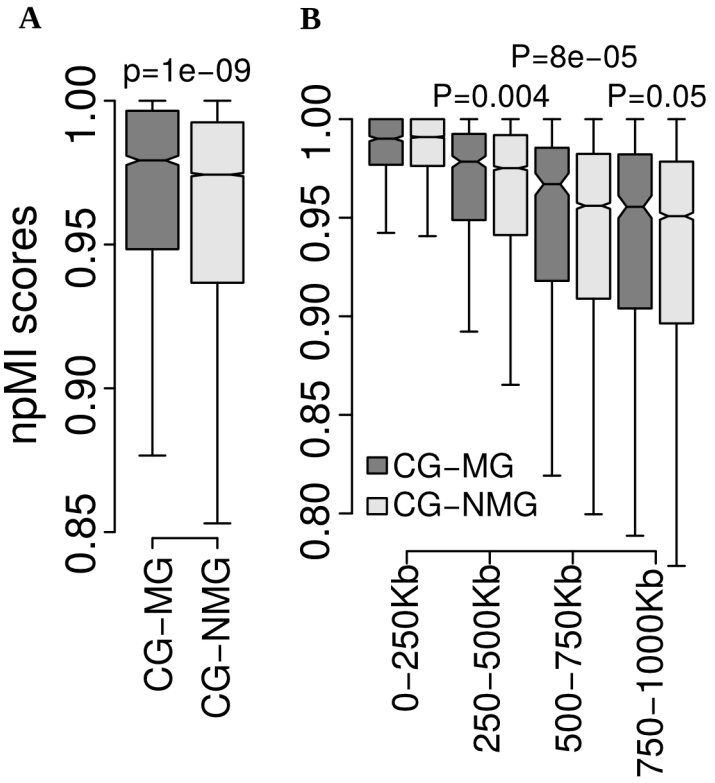


Figure 16: **Comparing co-occurrences of copy number alteration between cancer-metabolic and cancer-nonmetabolic gene pairs. (A)** Comparison of co-occurrences of copy number alterations (measured as information content, *npMI*) between all CG-MG and CG-NMG pairs from 19 cancer types. **(B)** Binned distance analysis for co-alterations between CG-MG and CG-NMG pairs separated by similar distances up to 1Mb using a sliding window of 250 Kb. The difference in the first bin is not significant.

4.3 IDENTIFYING ONCOGENIC FUNCTIONALITY OF CO-ALTERED CANCER-METABOLIC GENE PAIRS

We set up the *iMetCG* analysis pipeline to prioritize cancer-metabolic gene pairs with the objective of identifying metabolic genes involved in crucial metabolic reprogramming processes leading to functional consequences in tumorigenesis. The *iMetCG* pipeline (workflow in Fig. 12) performs statistical prioritization and exploits the concept of "guilt by association" for functional prioritization to identify metabolic genes involved in cancer metabolism. The pipeline was run on the 19 different cancer entities (Fig. 9 and Table 2) from The Cancer Genome Atlas (TCGA).

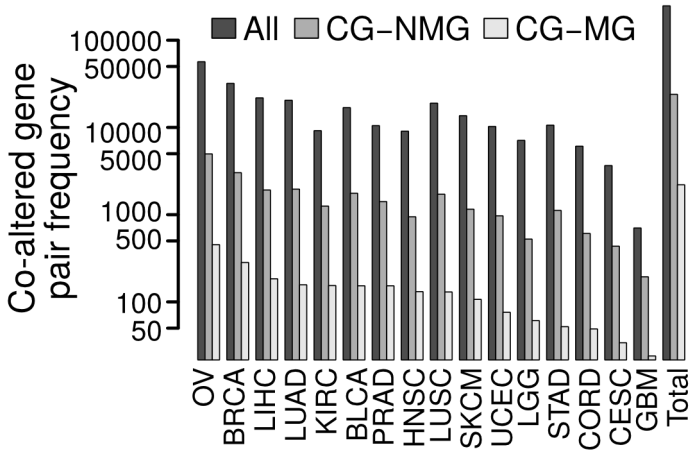


Figure 17: **Distribution frequency for co-altered gene pairs counts across tumors.** Co-altered gene pair counts in each cancer type analyzed, for all gene pairs (ALL), cancer-nonmetabolic (CG-NMG) gene pairs and cancer-metabolic (CG-MG) gene pairs The y-axis has been drawn to log scale and bars are sorted based on co-altered CG-MG gene pairs counts. These frequencies reflect only those that fulfilled our selection criteria at the statistical prioritization step.

Statistical Prioritization

Firstly, the information content was calculated for every gene pair in the 22 autosomes using robust normalized pointwise mutual information (*RnpMI*) as a statistical measure of its copy number co-alteration (*RnpMI* score distribution for all cancer types is given in Appendix Fig. 28). The *RnpMI* scores positively correlated with absolute co-occurrence probabilities (Appendix Fig.29). We used stringent *RnpMI* score filtering criteria to select distinctively co-altered gene pairs from each cancer type: (1) >99% quantile of the score distribution and (2) an absolute co-occurrence probability >5% using a relaxed gene pair distance of <5Mb. The frequency counts for these statistically prioritized gene pairs across all cancers is shown in Fig. 17. We selected only the cancer-metabolic gene pairs from this unbiased list of total significantly co-altered gene pairs from each cancer type for functional prioritization.

Functional Prioritization

Co-expression analysis

Statistically prioritized gene pairs were functionally prioritized by associating gene expression profiles from the matched cancer type to their copy number status. For each cancer type, we selected cancer-metabolic gene pairs for which metabolic

gene expression differed significantly ($p \leq 0.05$ with Benjamini-Hochberg correction for multiple testing) between all tumor samples harboring or lacking the copy number alteration (in both metabolic and cancer gene) relevant for the cancer-metabolic gene pair. None of the gene pairs reached our selection threshold from acute myeloid leukemia, papillary thyroid carcinoma or papillary kidney carcinoma, and these cancers were not considered in further steps. A unique set of 243 metabolic genes, irrespective of their co-alteration partners, was generated, for which 201 were associated with copy number amplifications and 43 were associated with deletions. *WHSC1L1* was implicated in both SCNA types. Interestingly, co-amplification

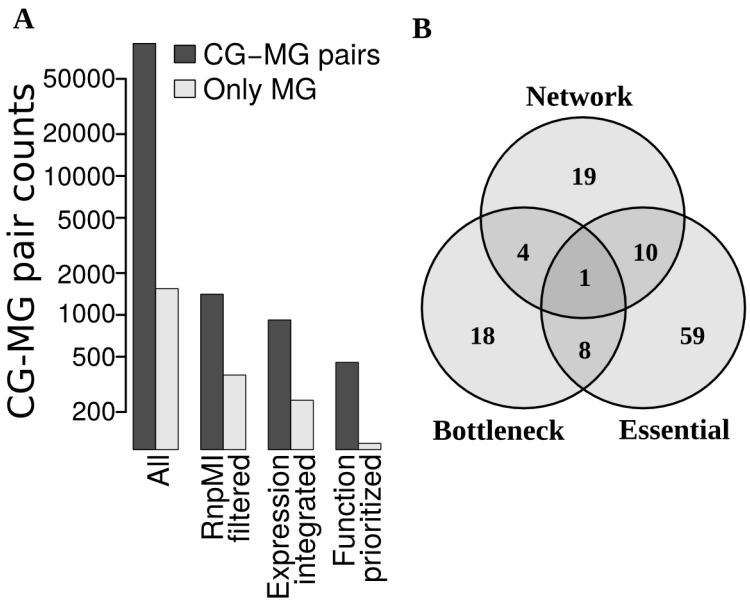


Figure 18: **Prioritization process of CG-MG pairs in *iMetCG* analysis pipeline**. (A) Cancer-metabolic gene pair counts at various filtering and prioritization steps in the *iMetCG* analysis pipeline. (B) Distribution of putative metabolic cancer genes in different *a priori* defined functionally relevant gene lists.

events were more common than co-deletion events in these identified pairs, suggesting that the cancer has additional benefits from retaining genes and reprogramming its expression than losing genes in deletion events. An overview of the pathways targeted with these co-altered metabolic genes in the global human metabolic map is shown in Fig. 24. We refer to these identified pairs as "prioritized and co-expressed cancer-metabolic gene pairs" and refer to the unique set of individual metabolic genes as the "core metabolic gene set".

Integration of a priori defined gene lists and network analysis

The *iMetCG* pipeline further filtered the prioritized and co-expressed cancer-metabolic gene pairs by integrating metabolic functionality features with the data. We hypothesize that metabolic genes important for normal metabolism, as defined by their essentiality, rate-limiting property and high connectivity in the metabolic network, would also exert a functional phenotype when co-altered in the cancer. This relationship to existing metabolic functionality was analyzed by identifying genes from the core metabolic gene set belonging to an *a priori* defined gene list of essential and bottleneck metabolic genes. We also employed a gene-centric global human metabolic network to identify meta-bolic genes whose nearest neighbors (N$\geq$5) in the network were enriched for differentially expressed genes in expression profiling datasets from tumors with copy number alterations in the proximal cancer genes. This resulted in a unique set of 119 metabolic genes, 95 of which were associated with amplifications and 25 of which were associated with deletions. *WHSC1L1* was again implicated in both SCNA types, and involved 34 genes identified from the network analysis, 78 essential genes and 31 bottleneck metabolic genes (Fig. 18B). We term these "putative metabolic cancer genes", identified after selection using metabolic functionality and multiple levels of prioritization. Taken together, our *iMetCG* analysis started with 89,889 cancer-metabolic gene pairs, and identified 119 putative metabolic cancer genes. A total of 919 pairs were obtained following statistical prioritization and co-expression analysis, and a final set of 528 gene pairs were identified upon functional prioritization (Fig. 18A). These identified gene pairs provide an elaborate set of novel and known metabolic cancer genes very likely to be involved in metabolic reprogramming of tumor cells, of which, we give some intriguing examples of in the next sections.

iMetCG analysis pipeline identifies significantly co-altered cancer-metabolic gene pairs that have a functional effect in tumors

4.4 IDENTIFICATION OF SYNERGISTIC METABOLIC CLUSTERS

Data were assembled into a binary table that counted only the presence or absence of a co-altered gene pair in a cancer type. Clustering analysis using the binary distance and Ward's linkage of pairs identified in at least 2 out of 16 cancer types clustered the 405 pairs into well defined metabolic modules suggesting a common rewiring mechanism operating in related cancer types (Fig. 19, the complete heatmap with gene pair information in Appendix Fig. 30). These prioritized and co-expressed cancer-metabolic gene pairs include many well known tumor suppressors and oncogenes that were identified in

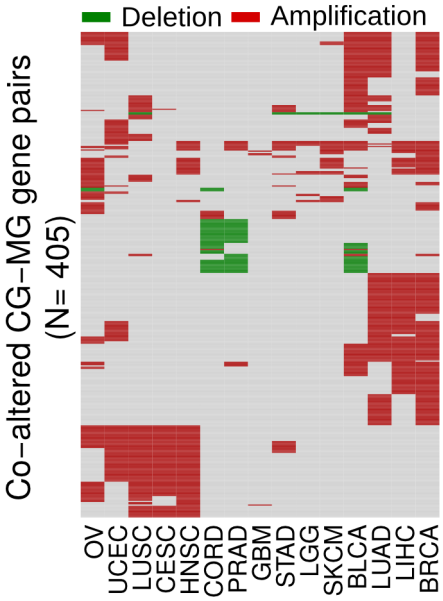


Figure 19: **Heatmap of the identified cancer-metabolic gene pairs co-altered across cancers.** A binary (alteration present: red/green, alteration absent: gray) heatmap for all cancer-metabolic gene pairs identified in at least two cancer types using the *iMetCG* analysis pipeline.

Table 3: **Representative examples of copy number co-altered cancer-metabolic gene pairs which are well-characterized tumor suppressor and oncogenes**[†].

Known oncogenes or tumor suppressors	**Proximal co-altered metabolic genes**
CDKN2A/2B	*MTAP*
MYC	*SQLE, NDUFB9*
RECQL4	*DGAT1 ,CYC1, GPAA1, TSTA3, OPLAH, PYCRL*
SOX2	*ABCC5, PIK3CA, ALG3, MCCC1, POLR2H, EHHADH, B3GNT5, DGKG, NDUFB5*
EGFR	*PSPH*
CCNE1	*UQCRFS1*
FGFR1	*WHSC1L1*
SDHC	*B4GALT3, NDUFS2, PPOX, PIGM*

[†]Examples taken from from Fig. 19

recent pan-cancer SCNA studies (Kim et al., 2013; Zack et al., 2013), suggesting that well known SCNA events occurring in the cancer genome often target proximal metabolic genes (Table 3). Analysis of frequency distribution of these co-altered pairs across cancer types and the common chromosomes involved

in such events revealed that most of these co-alteration events arise from few chromosomes suggesting an intrinsic genome dependent mechanism at play, which results in higher rates of SCNA co-alteration susceptibility for these chromosomes irrespective of the cancer type (Fig. 20). We observed a high redundancy in the co-altered partners shared by different cancers from metabolic-cancer gene pairs separated by <1 Mb. We interpret these data to mean that a single cancer gene can have multiple metabolic genes in its proximity and *vice verse*, or that genomic clusters of multiple metabolic and cancer genes may exist. Such genomic regions targeted by SCNA would create mul-

Many well known cancer causing genes are often co-altered with crucial metabolic genes in cancer

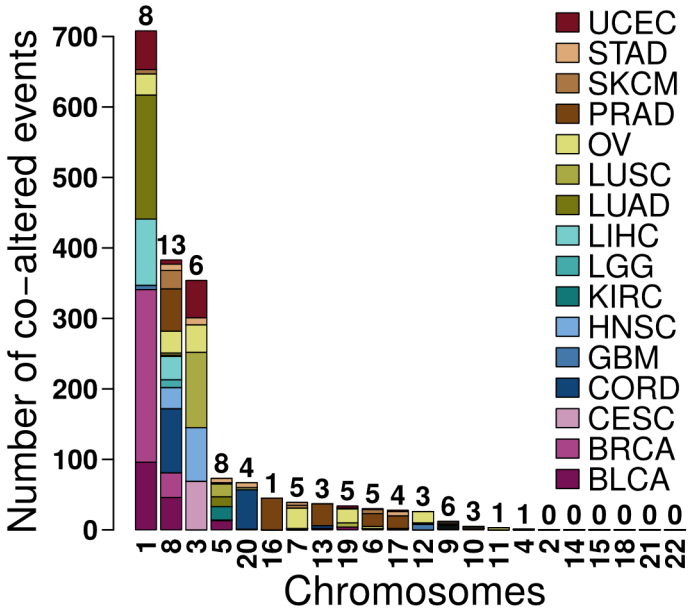


Figure 20: **Distribution of cancer-metabolic gene pairs in each cancer and chromosome.** Frequency distributions of co-altered CG-MG gene pairs after statistical prioritization and co-expression analysis in each chromosome. The data is from all cancer types (for full names of the cancers, refer Table 2). Numbers on top of each bar indicate the number of cancers in which the respective chromosome has at least one co-altered gene pair. The bars are sorted according to the number of co-altered CG-MG pairs.

tiple co-alteration partners for a single cancer or metabolic gene. We identified most commonly altered metabolic and cancer genes with different partners across cancers, thus, identifying susceptible loci in the genome where the chances of a metabolic gene being co-altered with a cancer gene is higher simply due to its proximity to multiple cancer genes. We identified two such SCAN-susceptible loci on chromosome 1, between q21.2 and q23.3 for amplification, and for deletion between p22 and p21.3 on chromosome 8. SCNA events in these loci that are densely populated with cancer and metabolic genes, presumably targeting a driver gene, consequently perturb multiple

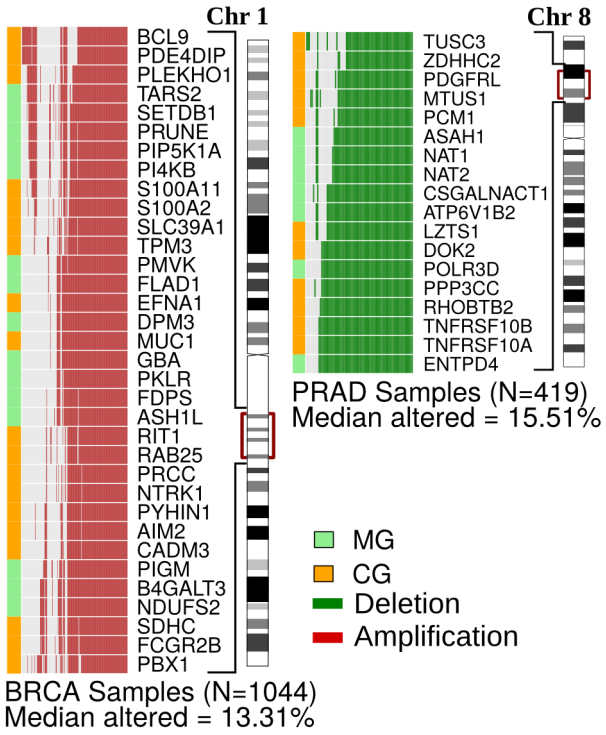


Figure 21: **Proximal metabolic and cancer gene rich genomic clusters coaltered in tumors.** Genomic regions in chromosome arms 1p and 8q, harboring cancer and metabolic gene rich clusters. These loci are affected in multiple cancer types, however only those cancers (breast - BRCA and prostate - PRAD) with the highest median alterations are shown.

metabolic genes and are observed in different cancer types (Fig. 21). We further used the core metabolic gene set of 113 genes affected in at least 2 cancers in a second clustering analysis to identify metabolic genes altered across cancer types irrespective of its co-alteration partner. (Fig. 22 shows metabolic genes altered in at least 25% of cancer types studied. The heatmap for all metabolic genes is given in Appendix Fig. 31). The common core metabolic genes identified here were perturbed synergistically across multiple tumor types again suggesting a common rewiring mechanism. We used these to classify tissues and developmental origins of the tumor types in the next section.

Many metabolic genes are commonly co-altered across cancer types either with the same or different cancer gene partner

4.5 TUMOR CLASSIFICATION BASED ON TISSUE AND DEVELOPMENTAL ORIGINS

Cancers are traditionally classified by histology and/or developmental origins to best define fitting diagnostics and treatment protocols (Berman, 2004) (also see, International Classification of Diseases for Oncology, `http://codes.iarc.fr/home`). It is increasingly acknowledged, however, that similar cancer

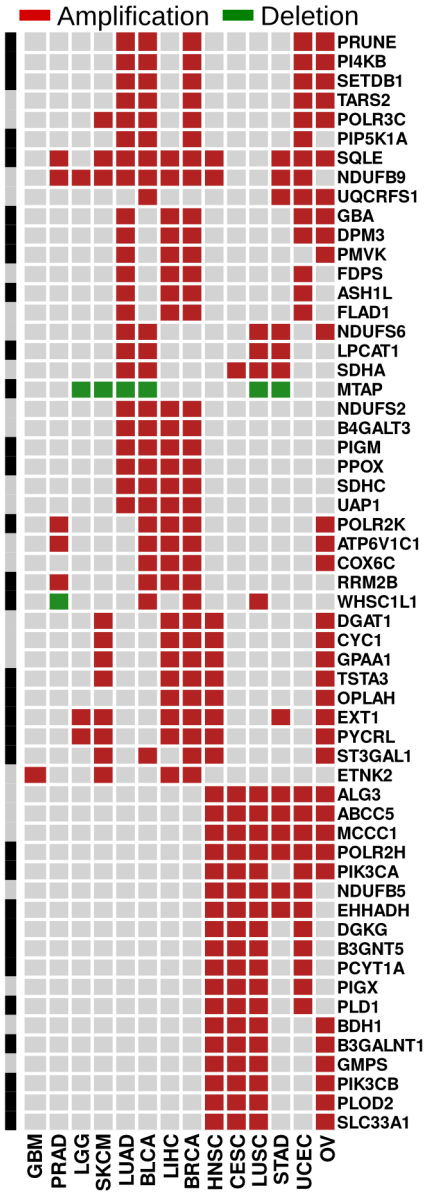


Figure 22: **Commonly co-altered metabolic genes across cancer types.** The *iMetCG* identified core metabolic gene set (for easier visualization, showing only those observed in at least 4 out of 16 cancers) across multiple cancer types after statistical prioritization and co-expression analysis, the color bar on the left shows the putative metabolic cancer genes (in black) identified in later steps.

types, a single type or even single tumors harbor vast molecular heterogeneity. Recent pan-cancer efforts have identified genetic commonalities across cancers suggesting core mechanisms for carcinogenesis. These genetic similarities stratify cancers into their tissues of origin (Network et al., 2013; Kim et al., 2013; Hoadley et al., 2014). Using the core metabolic gene set altered in at least 2 cancer types identified by the *iMetCG* pipeline, we

were able to group tumors into clusters that fit their previously described developmental or tissue origins and further identify novel clusters (Fig. 23). Since we analyzed more TCGA cancer types than prior studies, we also added novel information to existing tumor groups. We re-identified groups having lung adenocarcinomas, urothelial bladder cancers and head and neck as well as lung squamous cell carcinomas, as reported elsewhere (Hoadley et al., 2014). Our analyses also clustered cervical cell cancers, a squamous cell carcinoma not included in the previous study, with other squamous cell cancers. Our analyses also classified tumors into mesenchymal-like clusters containing ovarian cancers and endometrial carcinomas, which have mesodermal origins, and stomach adenocarcinomas. Stomach

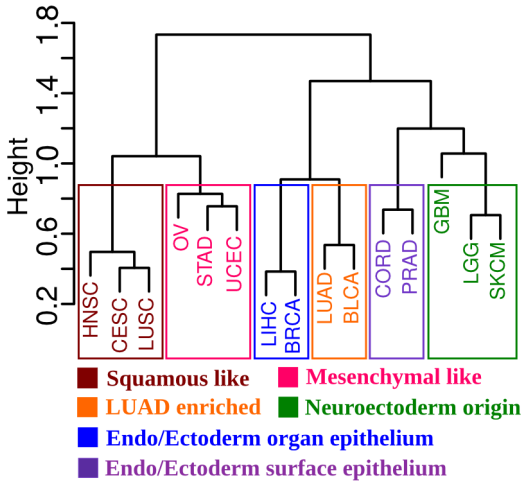


Figure 23: **Tumor classification.** Classification of tumors based on its developmental and tissue of origins using the core metabolic gene set.

adenocarcinoma tissues, though not of mesodermal lineage, are known to undergo epithelial-mesenchymal transition during tumorigenesis, (Katoh, 2005) which may be why they cluster with this group. Cancers from ecto- and endodermal lineages clustered into three groups: (1) from organ epithelium, including hepatocellular carcinomas and breast cancer, (2) from surface epithelium, including colorectal and prostate adenocarcinomas, and (3) of neuroectodermal origin, including glioblastomas, lower grade gliomas and cutaneous melanomas. Taken together, our results show that the *iMetCG* pipeline performs well in identifying a core subset of metabolic genes that can correctly classify tumors by its histological and developmental origins. This is remarkable considering that a rather small 113 metabolic gene signature is sufficient for this physiologically complex class separation.

iMetCG identified metabolic genes enable classification of tumors into distinct groups based on developmental or histological origins

4.6 PERTURBATIONS TARGETING METABOLIC PATHWAYS

To identify biochemical pathways affected across cancer types, we mapped the core metabolic gene set identified by *iMetCG* into a global human metabolic network based on pathway definitions using the Kyoto Encyclopedia of Genes and Genomes (KEGG `http://www.genome.jp/kegg/`). We observed that major metabolic processes relevant for cancer, including metabolic pathways for nucleotides, lipids/glycans, energy/carbohydrates and amino acids, were targeted by genes altered in various cancer types (Fig. 24). We focused our interpretations mostly

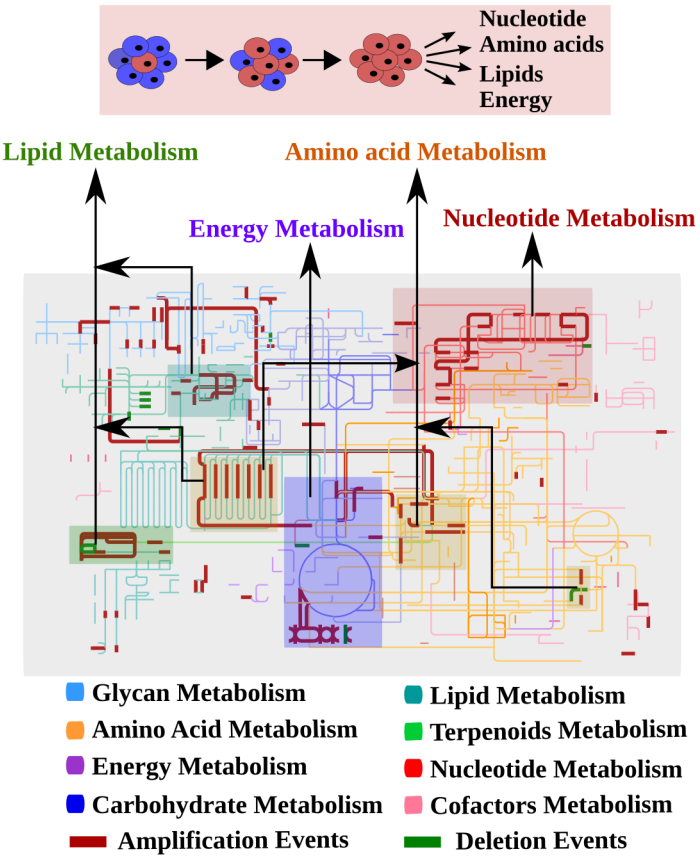


Figure 24: **Cancer-metabolic gene pair copy number co-alteration induced perturbations in the global human metabolic network.** Major cancer specific metabolic pathways targeted by the core metabolic gene set identified by the *iMetCG* pipeline are mapped onto the global KEGG human metabolic network.

on co-amplification events, unless otherwise stated, since these were most commonly observed in our analysis. Genes coding for enzymes in nucleotide metabolism (*NME7, GMPS, ADSS, RRM2B, UCKL1* and *UCK2*), which convert the monophosphate forms of purines or pyrimidines and their deoxy-counterparts to their respective di/tri phosphates, were identified using our

core metabolic gene set as being affected across cancer types as were the RNA polymerase subunits, *POLR3C, POLR2K* and *POLR2H*. The *MTAP* gene was co-deleted with the *CDKN2A/2B* tumor suppressor gene in 6 of 16 cancers. *MTAP* is crucial for salvaging adenine and methionine in the cell. This co-deletion is an example of a previously described event in multiple cancer types, validating our methodology to identify meaningful alterations for cancer (Fig. 26) (Bertino et al., 2014). Genes coding for critical enzymes from glycolysis, tricarboxylic cycle and fructose/mannose metabolism, such as *PFKFB2, PKLR, SDHC* and *SDHA*, were identified as altered across cancer types. Interestingly, we also identified enzymatic reactions known to produce intermediate metabolites in glycolysis involved in the allosteric control of glycolytic flux. The bi-functional enzyme, *PFKFB2*, that converts fructose-6-phosphate to fructose-2,6-bisphosphate, has been shown to tightly control this reversible reaction. The level of fructose-2,6-bisphosphate further controls the rate of glycolytic flux by regulating the conversion of fructose-6-phosphate to fructose-1,6-bisphosphate, while the latter is an allosteric regulator of the terminal glycolytic reaction step, where phosphoenolpyruvate is converted to pyruvate (Ros and Schulze, 2013). In addition to carbohydrate metabolism, we identified many genes involved in oxidative phosphorylation, particularly of the NADH dehydrogenase complex. Evidence is accumulating that oxidative phosphorylation is active and functional in many cancers, and tightly regulated by tumor suppressors and oncogenes (Jose et al., 2011). We identified *NDUFB2, NDUFA9, NDUFB9, NDUFS2, NDUFS6* and *NDUFB5*, all coding for subunits of the NADH dehydrogenase complex, as being co-amplified with cancer genes and belong to different chromosomes suggesting a common mechanism of co-alteration based metabolic remodeling for this complex. We also identified the genes for the cytochrome c oxidase/reductases complex, *CYC1, UQCRFS1, UQCRB* and *COX6C*, and genes for the ATPase complex, *ATP6V1C1* and *ATP5E*, as being altered across cancer types (Fig. 25). Glycerolipids are the basis for phospholipids, which are essential components of cell membranes. Phosphatidic acid has also been shown to tightly regulate *mTOR* function (Foster, 2007), hence, energy metabolism. Our analysis also identified genes involved in glycero(phospho)lipid metabolism, including *AGPAT6, PLD1, DGKG, ETNK2* and *PCYT1A*, as being affected across multiple cancers. These enzymes assist production of phosphatidic acid, which is a vital precursor for almost all glycerolipids (Athenstaedt and Daum, 1999). *HMGCS1, PMVK, ID1* and *FDPS* were also affected, which are involved in the

iMetCG identified metabolic genes belong to pathways related to biomass and energy production

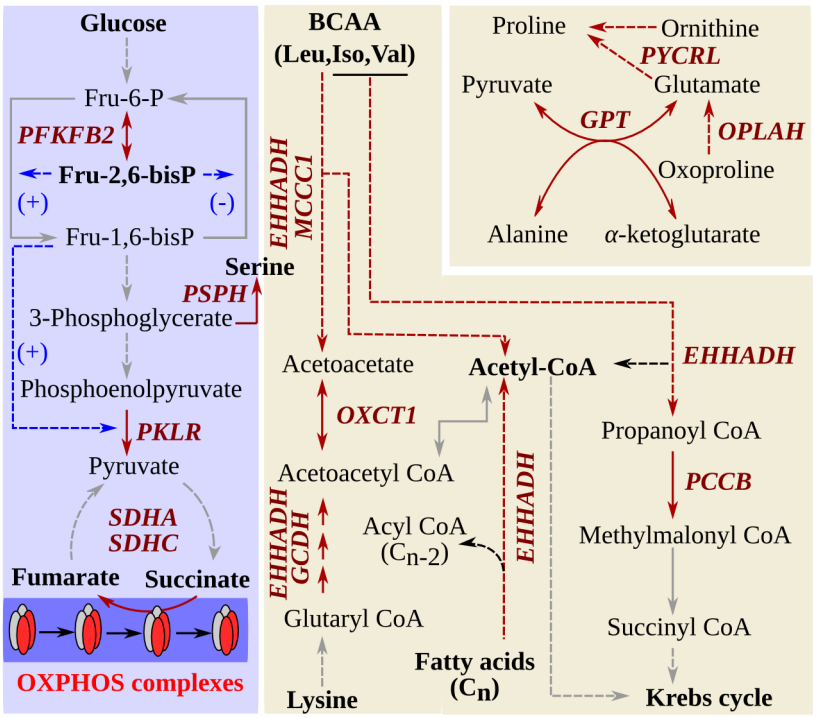


Figure 25: **Representative examples of co-altered metabolic genes targeting energy and amino acid metabolism.** Bold red arrows (single step reactions) or dashed red arrows (involving multiple reactions steps) represent genes identified by *iMetCG*. Gray arrows (both bold and dashed) are used to represent the context and continuity of the pathway (MG were not identified for these reactions). Blue dashed lines indicate the regulatory effect of the metabolite involved, (+) for positive and (–) for negative effects.

mevalonate pathway that ultimately produces farnesyl phosphate, an important precursor of many anabolic processes. The *SQLE* gene was most frequently altered across cancers (in 10/16 cancers), which encodes the rate limiting enzyme in steroid biosynthesis, suggesting an important role of steroid anabolism across tumors types (Fig. 26). In amino acid metabolism, several genes for branched chain amino acid metabolism were affected that are also common to fatty acid metabolism, including *EHHADH*, *MCCC1*, *PCCB* and *OXCT1*. Other affected genes from the amino acid metabolic pathway were *PSPH* and *PYCRL* from serine and proline metabolism, respectively, and *GPT* and *OPLAH* from glutaminolysis (Fig. 25). In total, the *iMetCG* pipeline performs well in identifying both known and novel metabolic processes and its genes crucial for cancer metabolism. Many of these identified genes have been shown to be important in cancer while the novel genes highlight interesting and unknown aspects of metabolic network remodeling.

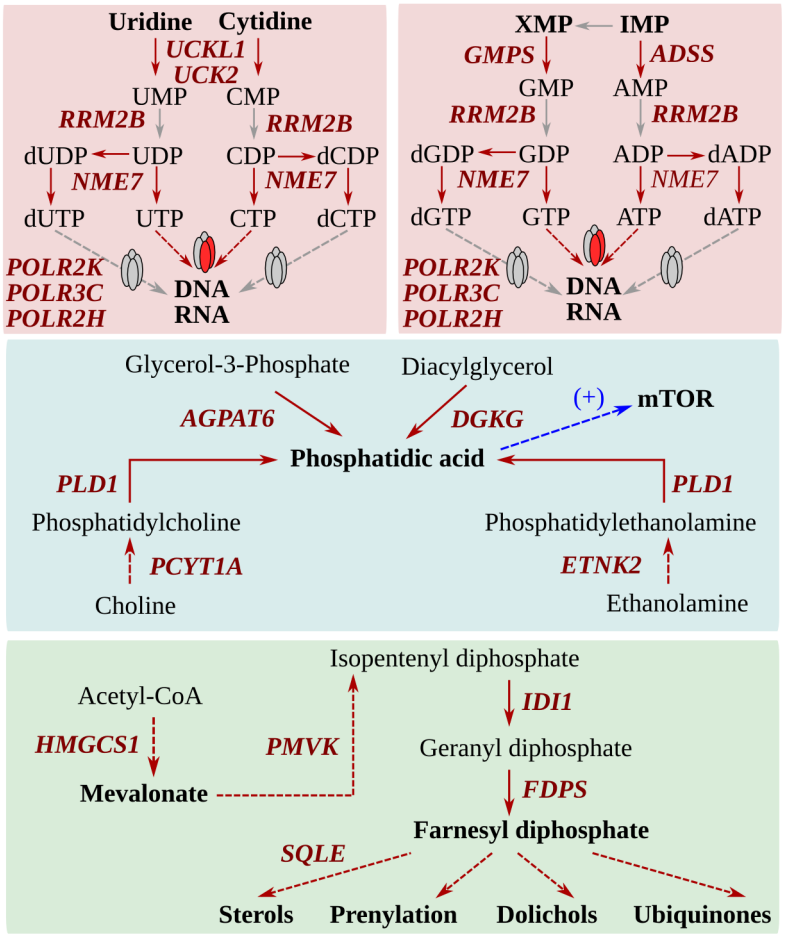


Figure 26: **Representative examples of co-altered metabolic genes targeting nucleotide and lipid metabolism.** Annotation and representational details is same as in Fig. 25 above.

4.7 PUTATIVE METABOLIC CANCER GENES SHARE SIMILAR PROPERTIES WITH CANCER GENES

4.7.1 *Putative metabolic cancer genes are enriched for known cancer genes and affect cancer survival*

We compared putative metabolic cancer genes identified in *iMet-CG* with a list of known cancer-causing metabolic genes then performed an enrichment test to determine if the former was overrepresented in this list. Indeed, we observed that the relative percentage of putative metabolic cancer genes was significantly higher than non-metabolic cancer genes (p=0.006, using a one-sided Fisher's Exact test, Fig. 27A). We also evaluated the effect of putative metabolic cancer genes on cancer survival using a log-rank test for every metabolic gene having high-level copy number alterations (amplifications or deletions) in at least

5% of the samples. This analysis computed the survival difference between cancer samples harboring or lacking copy number alterations. Strikingly, the relative percentage of survival events involving a putative metabolic cancer gene was significantly higher than survival events involving other metabolic genes (p=0.002, using a one-sided Fisher's Exact test, Fig. 27B), suggesting that these genes are a crucial determinant of cancer survival across the investigated cancer types.

4.7.2 *Putative metabolic cancer genes have higher network connectivity*

In the context of a human metabolic network, like any other cellular network, disrupting highly connected genes or hubs can have a drastic impact on function (Jeong et al., 2001). Along these lines, we propose that cancer-driving metabolic gene candidates should be highly connected compared with other metabolic genes in the network, such that cancer cells can reprogram metabolism more effectively. We estimated gene connectivity for each metabolic gene using our gene-centric metabolic network, and compared network connectivity of putative metabolic cancer genes with other metabolic genes. We observed that the average connectivity of these *iMetCG*-identified genes was significantly higher than other metabolic genes in the network (p=0.05, Fig. 27C), suggesting their disruption is more likely to affect central nodes in the metabolic network and, hence, more likely to remodel the cancer metabolic landscape.

4.7.3 *Putative metabolic cancer genes and their evolutionary context*

Cancer genes are highly conserved, while different isoforms can be important determinants of cancer progression and malignancy (Thomas et al., 2003; Oltean and Bates, 2014). A well-known example is the gene isoform *PKM2*, which plays a crucial role in anabolic metabolism promoting aerobic glycolysis and tumor growth (Vander Heiden et al., 2010). We investigated whether *iMetCG*-prioritized putative metabolic cancer genes share similar features of evolutionary conservation, namely selection pressure and evolutionary rate, and/or similar transcript isoform diversities. The ratio of non-synonymous (dN) to synonymous (dS) substitutions in orthologous genes of distant species gives a robust measure for selection pressure. A $\frac{dN}{dS} < 1$ implies purifying selection (higher conservation), while a $\frac{dN}{dS} > 1$ suggests positive selection and lower conservation (Kimura, 1977). We used the *GRCh38* Ensemble release for human and

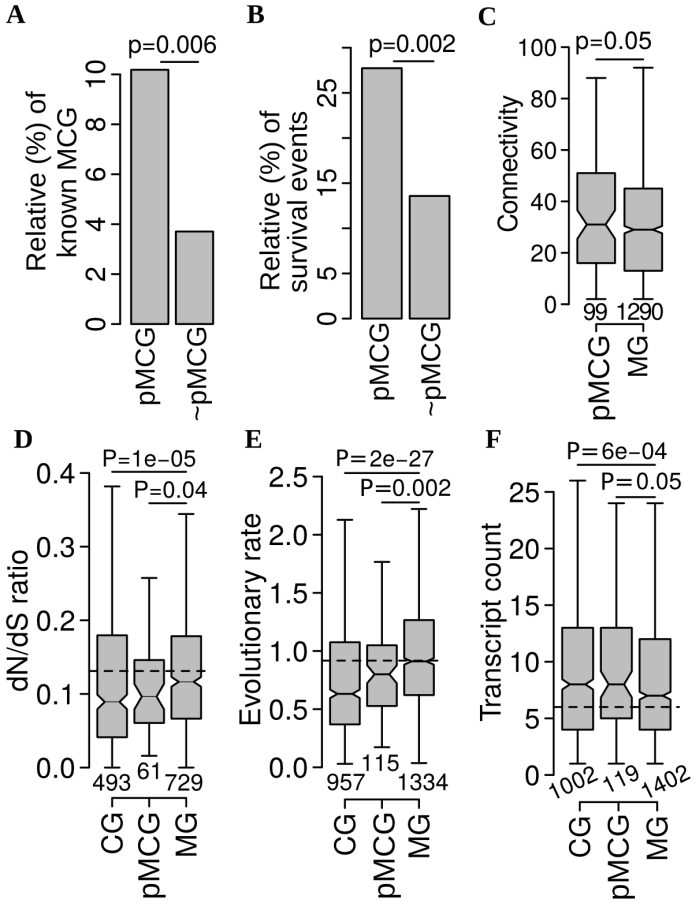


Figure 27: **Known cancer genes and the identified putative metabolic cancer genes share similar features.** Enrichment of putative metabolic cancer genes (pMCG) – (**A**) in the list of known metabolic cancer genes (MCG) (**B**) among other metabolic genes affecting cancer survival. ~pMCG denotes non-pMCG, i.e those metabolic genes that were not identified in our analysis as cancer causing genes. (**C**) Comparison of network connectivity (i.e. the number of nearest neighbors in the gene centric human metabolic network) between putative metabolic cancer genes and all other metabolic genes, the numbers below the boxplot represent the sample sizes. Comparison of genomic parameters – (**D**) selection pressure (**E**) evolutionary rates and (**F**) isoform diversity among cancer genes, metabolic genes and putative metabolic cancer genes. Dashed black lines represent the median values of each feature analyzed for all genes. Sample sizes are shown under respective boxplots.

GRCm38.p2 for mouse in Biomart to obtain $\frac{dN}{dS}$ ratios for human-mouse orthologous genes. It has been recently shown that cancer and metabolic genes are evolutionarily more conserved and under stronger purifying selection than other genes (Cheng et al., 2014). We made a similar observation while comparing the distributions of $\frac{dN}{dS}$ ratios for cancer and metabolic genes (p=1e-05) (Fig. 27D). Interestingly, we observed that *iMetCG*-identified putative metabolic cancer genes share similar median $\frac{dN}{dS}$ ratios with cancer genes, which is lower compared to the re-

maining metabolic genes (p=0.04). We also compared the evolutionary rates among cancer, metabolic and putative metabolic cancer genes using data calculated elsewhere (Bezginov et al., 2013). In concordance to our calculated $\frac{dN}{dS}$ ratios, we observed that putative metabolic cancer genes have distinctively lower evolutionary rates than other metabolic genes (p=0.002, Fig. 27E). We also observed that isoform diversity of putative metabolic cancer genes shared similar isoform counts with cancer genes but have higher than average isoform count compared to the remaining metabolic genes in the genome (p= 0.05, Fig. 27F). Based on various cancer relevant features we analyzed, the *iMetCG*-identified putative metabolic cancer genes were noted to be very similar to known cancer genes while being distinctively different from remaining metabolic genes. Since feature conservation generally translates into functional relatedness in biology, this goes in line with our hypothesis that these genes are central to cancer cell metabolism.

"Guilty by association"- iMetCG predicted metabolic cancer genes share similar phenotypic, genetic and evolutionary properties with known cancer genes

5

CONCLUSIONS

[1]In this thesis it was hypothesized that, if genome evolution has preferentially positioned genes vital for cellular proliferation near cancer drivers in the genome or *vice versa*, then they (cancer causing and their proximal genes) would be more susceptible to copy number co-alterations events. Our hypothesis is based on recent findings that highlight the tumorigenic potential of genes closely positioned to cancer causing genes in the genome (Solimini et al., 2012). Furthermore, exploiting this proximity advantage to concomitantly affect these crucial nearby genes would present a highly beneficial strategy to cancer progression. In this study, we specifically focus on pairs of proximal metabolic and cancer causing genes and show how they are affected by copy number co-alterations across multiple tumor types. Moreover, we show that this linear proximity between oncogenes/tumor suppressor genes and metabolism specific genes in the chromosomes affects reprogramming of tumor metabolism. It was observed that cancer-metabolic gene pairs were significantly closer together in the genome compared to cancer-non-metabolic genes. Furthermore, we found across tumor types that the cancer-metabolic gene pairs were more often co-altered than cancer-nonmetabolic genes (Fig. 14,16).

This observed proximity between cancer-metabolic genes and their higher rates of copy number co-alterations suggested that these co-altered metabolic genes might affect tumor metabolism through its increased (amplifications) or decreased (deletions) gene dosage. It was hypothesized that the cumulative changes in gene dosage may play a major role in remodeling tumor metabolism. We propose that copy number co-altered metabolic genes individually might not yield an oncogenic effect but the cumulative effect of all these alterations together can support the cancer phenotype. This concept stems from the fact the cancer genomes harbor a large number of presumably neutral mutations that are considered to have no functional effects. However, recent studies have shown that there exists a vast number of oncogenes and tumor suppressor genes that exert their functional effect via triplosensitivity and haploinsufficiency respectively (Solimini et al., 2012; Davoli et al., 2013). Another interesting possibility is the "non-oncogenic addiction" of a cancer

1 Parts of this chapter has been taken from (Sharma et al., 2015)

cell on the co-altered metabolic genes identified in this study. Non-oncogenic addiction refers to the conditional dependency of a cancer cell on certain genes (for example, genes involved in stress response) which are however non-essential for a normal cell (Luo et al., 2009). These genes on its own cannot cause an oncogenic transformation of a normal cell. However, its function becomes essential for the survival of a cancer cell once it has transformed due to mutations in other driving oncogenes or tumor suppressor genes (Luo et al., 2009). Moreover, such genes are effective therapeutic targets as the survival of the cancer cell is dependent on it while it is dispensable for the normal cell (Raj et al., 2011; Riabinska et al., 2013). It is to be noted that these genes play important functional roles even in a normal cell, although such genes are not required or used given that the physiological and the microenvironment conditions of a normal cell is vastly different from a cancer cell (Luo et al., 2009). In the context of metabolism it has been suggested that cancer specific dependency on glutamine and serine metabolism can be consequences of non-oncogenic addiction (Galluzzi et al., 2013). Overall it was expected that, if our identified metabolic genes supported crucial nutrient requirements in tumors then these co-altered metabolic genes will prove indispensable for tumor growth and proliferation. Hence a subset of these co-altered metabolic genes may in fact be crucial co-driving cancer genes expressing its non-oncogenic yet cumulatively essential roles in tumorigenesis.

The *iMetCG* analysis pipeline was developed in order to identify such copy number co-altered metabolic genes and stringently prioritize them to elucidate enzymes playing an essential role in tumor metabolism (Fig. 12). The *iMetCG* approach was systematically applied to 19 different cancer types. Using mutual information based scores we identified distinctively co-altered metabolic genes. This was followed by expression analysis where the copy number alterations of metabolic genes were correlated with their expression levels. Additionally, using *a priori* biological knowledge, network analysis and several selection criteria we determined the functionality of co-altered metabolic genes in tumors. Recent pan-cancer SCNA studies on co-occurring copy number alterations have avoided looking into co-alteration events on the same chromosome because of difficulty in interpreting the results due to spatial gene proximity (Klijn et al., 2010; Zack et al., 2013; Kim et al., 2013). In order to overcome this restriction and identify metabolic genes that might play a functional role in tumor metabolism, several additional steps for functional prioritization in the context of

metabolism and cellular functionality was performed. This was done to distinguish the relevant co-occurring copy number altered cancer-metabolic gene pairs from artifacts. These selection criteria involved exploiting known *a priori* biological knowledge for functional prioritization by selecting metabolic genes based on their essentiality and if they occupied crucial positions in a metabolic pathway such that they could control the metabolic flux. It has been shown that metabolic genes that are coding for either the initial or terminal enzymes in a pathway are usually under tight transcriptional regulation and also control the flux passing through that pathway. These gatekeepers, if perturbed can disrupt the whole pathway flux distribution with severe consequences (Wessely et al., 2011). Furthermore, a gene centric metabolic network was constructed and co-altered metabolic genes were prioritized if their copy number changes also led to similar expression changes in the genes with which they were directly connected to in the network (Fig. 12).

Interestingly, many metabolic genes identified using the *iMetCG* analysis pipeline were similarly co-altered in multiple cancer types. This observed pan-cancer redundancy further supported the idea that they play an important role in tumor metabolism (Fig. 22). Metabolic remodeling is a crucial hallmark of cancer, a rapidly proliferating cancer cell requires more nucleotides, amino acids and lipids to maintain high biomass production. Consequently, to supplement these increased nutrient demands, energy metabolism within the cell must be high and needs to be adjusted to facilitate the changing tumor metabolic landscape (DeBerardinis et al., 2008). Moreover, these requirements are tightly regulated by signaling cascades and depends on tumor physiology (metastatic, angiogenic etc), location (tissue of origin) and microenvironment (hypoxia, pH etc) (Fig. 3). Indeed, copy number co-altered metabolic genes identified using the *iMetCG* framework belonged to many metabolic pathways indispensable to cancer proliferation, such as nucleotide, amino acid, energy and lipid metabolism. Also, many of the identified genes are already known to play a crucial role in tumor metabolism suggesting that the *iMetCG* pipeline can correctly identify crucial cancer relevant metabolic targets (Fig. 24). Interestingly, using the *iMetCG* identified core metabolic genes, defined as those that were co-altered in at least two of the cancer types analyzed, enabled successful grouping of cancers into their tissue and developmental origins. This also suggests a rather specific metabolic role of these enzymes for the particular subgroups of cancer types. This is remarkable considering that a small signature of 133 metabolic genes de-

rived from their co-alteration propensity with proximal cancer causing genes could reproduce this biologically complex histological classification (Fig. 23). Additionally, this observation may have a substantial clinical relevance, since a therapeutic agent inhibiting one of these metabolic gene could effectively be used in multiple cancers of similar origin and would presumably have comparable drug-induced effects. The final set of putative metabolic cancer genes was identified after *a priori* biological knowledge integration and these genes shared very similar characteristics with known cancer genes. Here, the concept of "guilt by association" was used where it was assumed that if the identified putative metabolic cancer genes and known cancer genes shared genotypic-phenotypic properties that typically define an oncogene or a tumor suppressor gene then it is likely that the metabolic gene is a yet undiscovered cancer causing gene. Indeed, it was observed that these genes shared similar evolutionary parameters (evolutionary rate and *dN/dS* ratios) and isoform diversity with known cancer genes. Additionally, based on these properties, the identified putative metabolic cancer genes were significantly different from other metabolic genes. Furthermore, these metabolic genes were significantly enriched in known cancer-causing metabolic genes, had high network connectivity and were associated with patient survival. Overall these indicators strongly suggest a potential tumorogenic role for these predicted metabolic cancer genes (Fig. 27).

In this study, a novel resource for cancer-associated metabolic genes was generated using integrative genomic based approach and robust computational and statistical methods on large cancer datasets. A novel mechanism governing tumor metabolic reprogramming across multiple tumors has been presented where cancer-associated metabolic genes were derived from linearly proximal cancer-metabolic gene pairs which were significantly affected by copy number co-alterations. To systematically identify all such gene pairs across tumor types we have developed the *iMetCG* analysis pipeline. These identified putative metabolic cancer genes most likely play a vital role in cancer progression and survival by serving the atypical metabolic needs of cancer cells. Furthermore, this resource of novel predicted cancer-affecting metabolic genes will be useful to the scientific community to further evaluate and elucidate detailed mechanisms and consequences of such individual co-alteration events.

6

APPENDIX

Table 4: Distribution statistics (in Kb) for gene pair distances less than 1Mb in all chromosomes.

Chr.	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	Size
1	0.55	213.9	461.5	474.7	727.7	1000	26770
2	0.067	198	431.7	454.8	697.8	999.8	9834
3	0.352	206.9	439.2	459.7	703	999.9	9927
4	0.718	211.4	446.3	465.9	710.5	999.7	4346
5	0.797	200.7	448.2	467.7	727	999.8	5614
6	0.383	197.6	435.4	456.4	700.5	1000	11856
7	0.271	209.9	448.3	463	704.1	999.9	7994
8	0.041	204.7	434.3	456.9	700.1	999.5	5239
9	0.871	186	407.3	440.5	681.2	999.6	9234
10	0.133	208.2	451	467.3	715.7	1000	5401
11	0.044	216.3	466.8	477.8	735.1	999.9	21960
12	0.01	214.6	452.1	472.1	724.9	999.9	13120
13	6.572	221.8	454.9	472.3	702.5	999.4	1449
14	0.588	179.3	399.8	435.8	676.7	999.9	6413
15	2.4	228.2	468.4	479.2	721.5	1000	5280
16	0.296	204.1	437.6	457.6	703.2	999.9	14568
17	0.093	212.9	447.1	468.4	718.4	1000	23014
18	4.956	209.3	418.7	454.1	692.7	996.8	1182
19	0.53	232.7	481.8	487.9	738.6	999.9	35591
20	1.737	209	447.5	466	710.9	999.7	6667
21	1.23	149.5	343.4	408.9	651.7	998.7	2206
22	2.41	220.1	470.6	475.7	720.3	999.9	5845

Table 5: Statistical and graph properties of the constructed gene-centric metabolic gene network.

	Min.	**1st Qu.**	**Median**	**Mean**	**3rd Qu.**	**Max.**
Degree	1	12	28	32.45	46	168
Eccentricity	1	7	7	7.377	8	12
Shortest paths	1	3	3	3.48	4	12
Betweeness (node)	0	0.5	269.4	1571	1655	38930
Betweeness (edge)	1	20.92	64.84	135.8	148.1	9347
Clique sizes	2	4	7	9.497	12	53
Transitivity (local)	0	0.528	0.713	0.709	0.977	1

Number of nodes	1389
Number of edges	22538
Maximal clique	750
Transitivity (global)	0.557

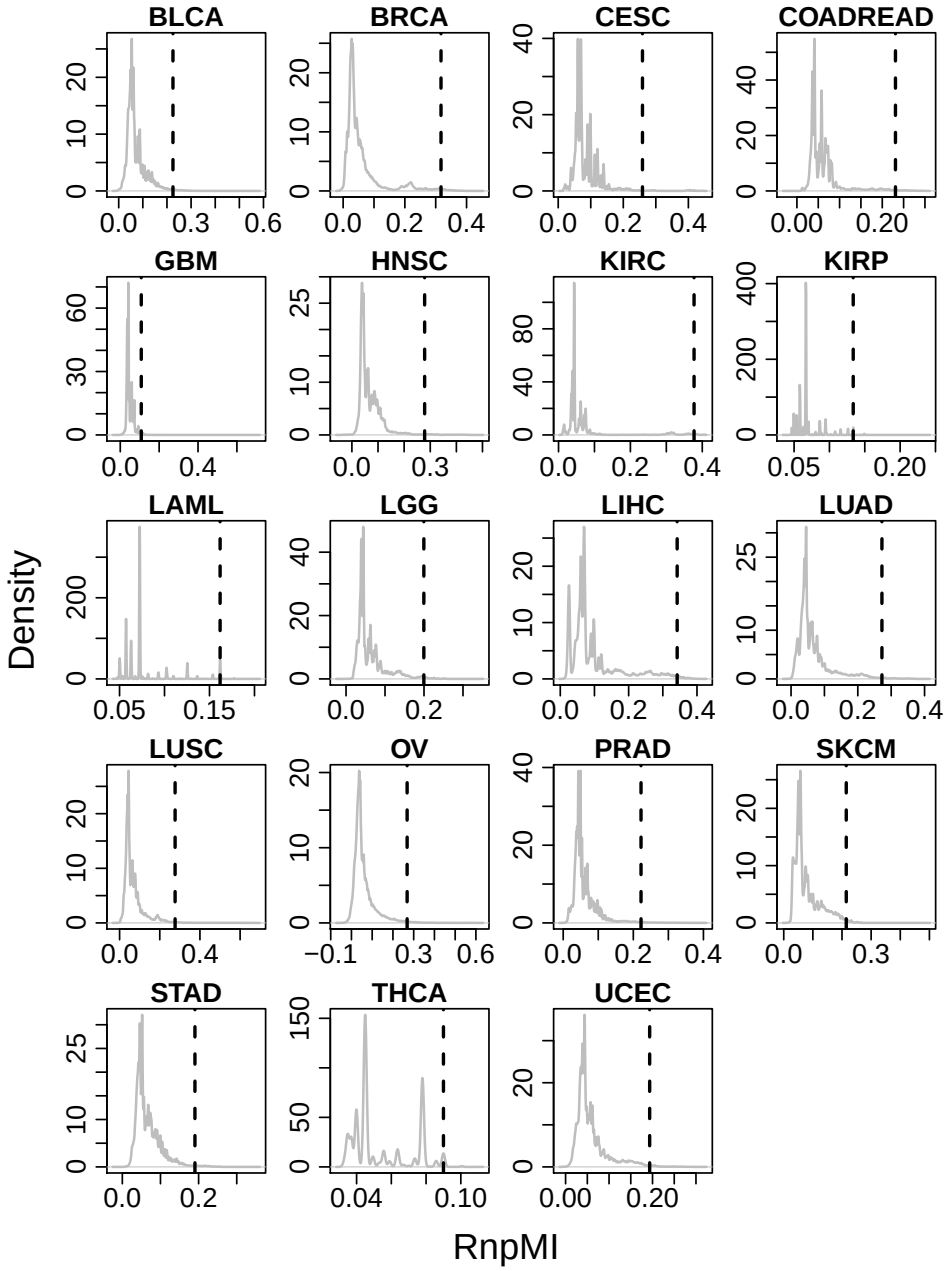


Figure 28: **Distribution of *RnpMI* scores across all cancer types.** The distribution of information content calculated from copy number co-alterations of all gene pairs separated by <5Mb. The 99% quantile of *RnpMI* score distribution was used as a threshold to prioritize the most affected gene pairs (shown as dashed black lines) in each investigated cancer type. For complete names of the cancers refer Table 2.

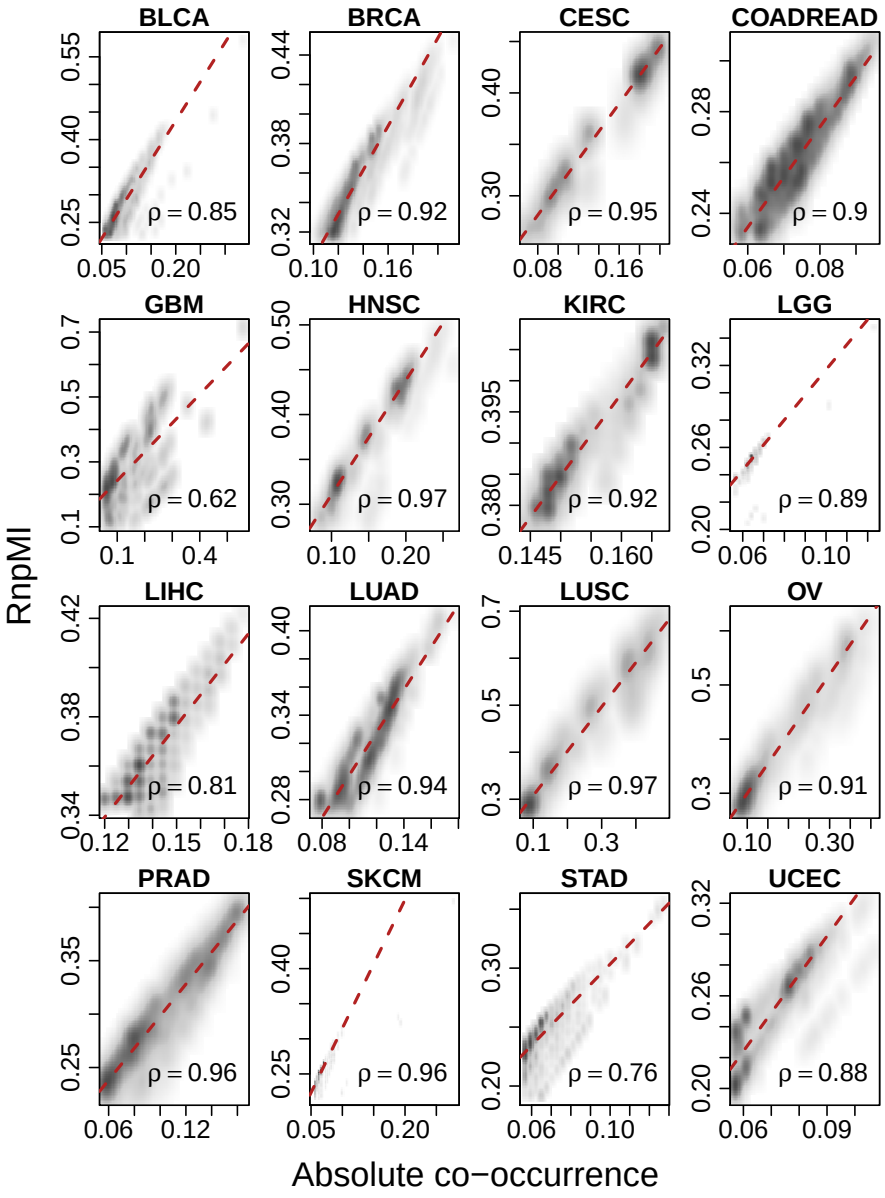
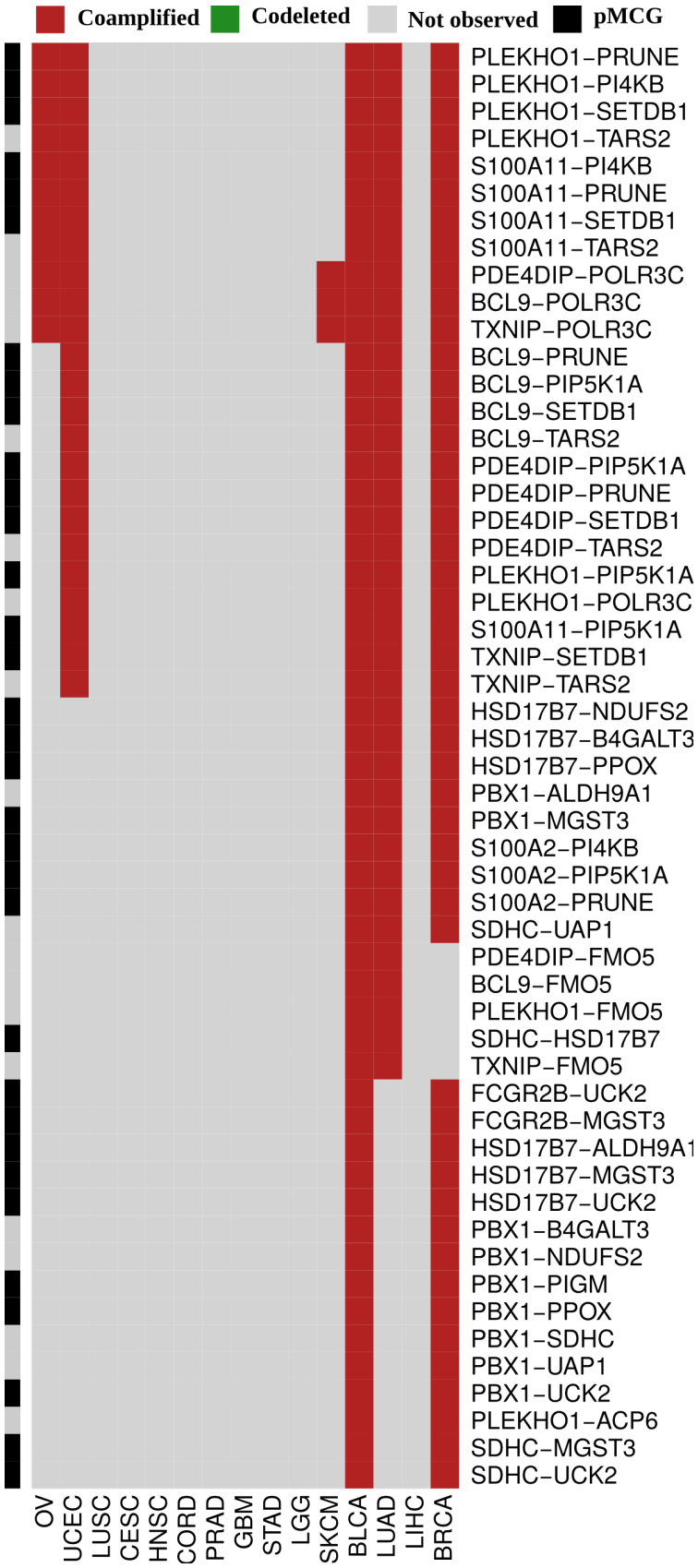
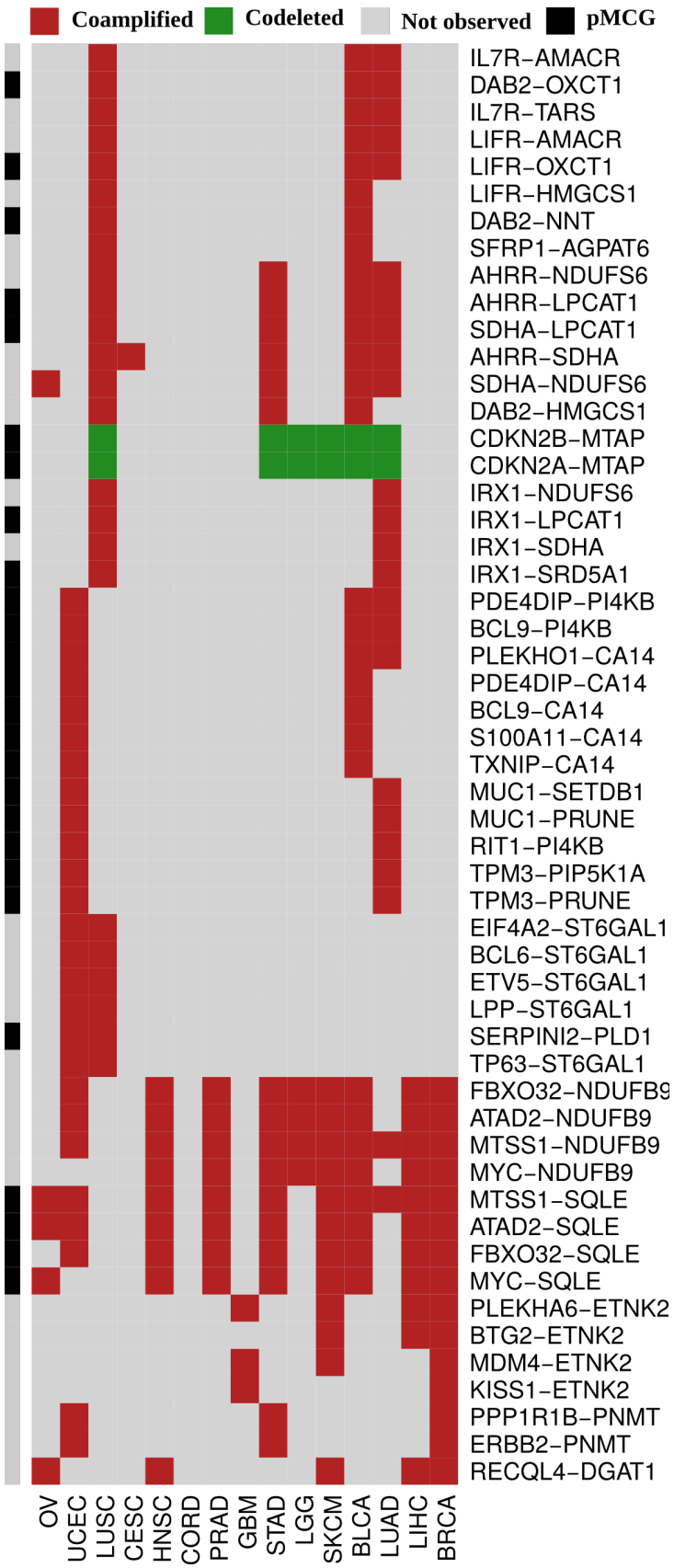


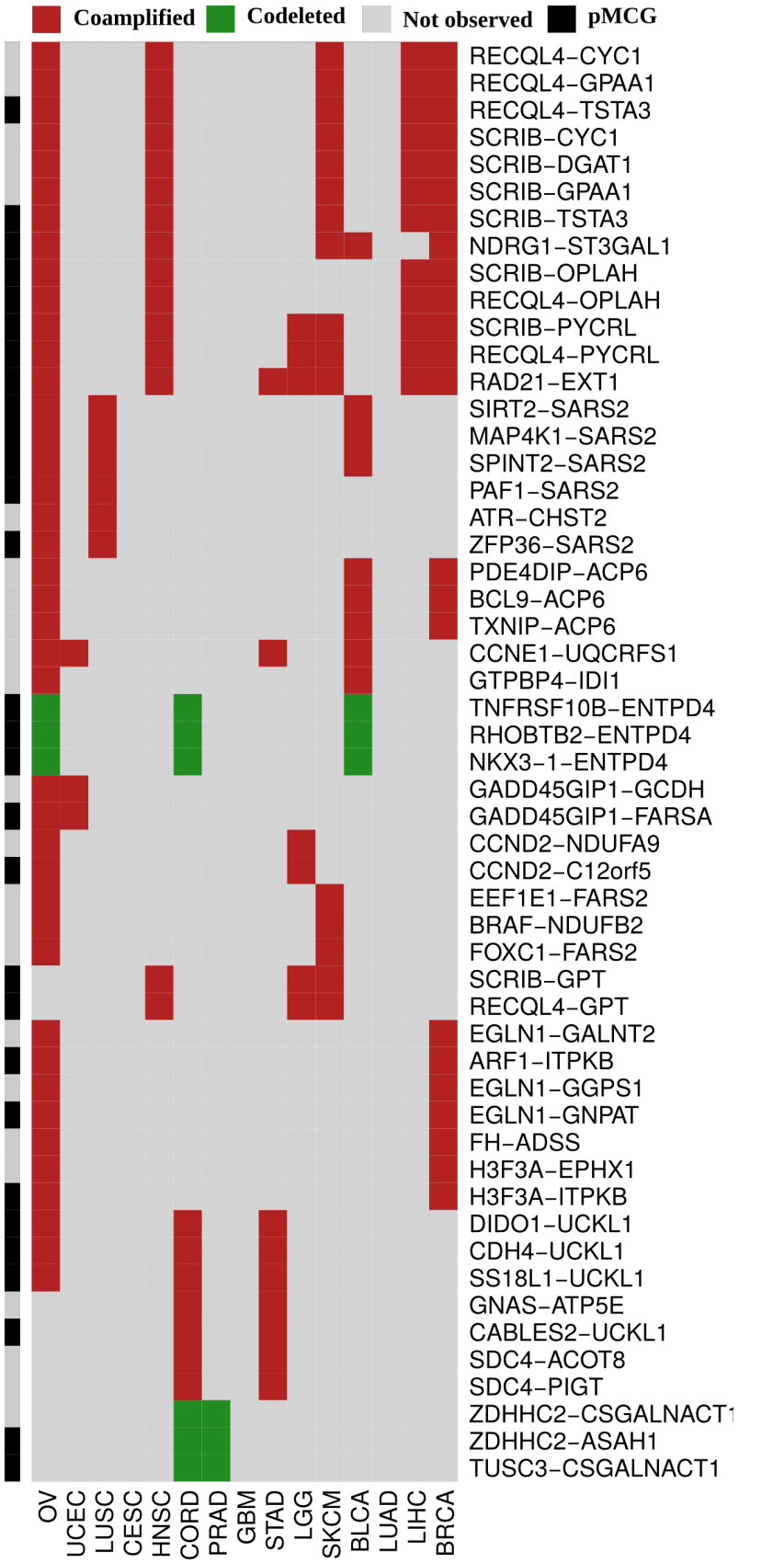
Figure 29: **Correlation between RnpMI scores and the absolute co-occurrence frequency.** A strong positive correlation (Spearman's correlation) was observed between the RnpMI based information content and the absolute percentage of co-altered gene pairs across all cancer types studied. The data is shown only for those gene pairs passing our selection criterion at the statistical prioritization step. The red dashed line represents the regression line. For full names of the cancers, refer Table2.


Coamplified
Codeleted
Not observed
pMCG
PLEKHO1–PRUNE
PLEKHO1–PI4KB
PLEKHO1–SETDB1
PLEKHO1–TARS2
S100A11–PI4KB
S100A11–PRUNE
S100A11–SETDB1
S100A11–TARS2
PDE4DIP–POLR3C
BCL9–POLR3C
TXNIP–POLR3C
BCL9–PRUNE
BCL9–PIP5K1A
BCL9–SETDB1
BCL9–TARS2
PDE4DIP–PIP5K1A
PDE4DIP–PRUNE
PDE4DIP–SETDB1
PDE4DIP–TARS2
PLEKHO1–PIP5K1A
PLEKHO1–POLR3C
S100A11–PIP5K1A
TXNIP–SETDB1
TXNIP–TARS2
HSD17B7–NDUFS2
HSD17B7–B4GALT3
HSD17B7–PPOX
PBX1–ALDH9A1
PBX1–MGST3
S100A2–PI4KB
S100A2–PIP5K1A
S100A2–PRUNE
SDHC–UAP1
PDE4DIP–FMO5
BCL9–FMO5
PLEKHO1–FMO5
SDHC–HSD17B7
TXNIP–FMO5
FCGR2B–UCK2
FCGR2B–MGST3
HSD17B7–ALDH9A1
HSD17B7–MGST3
HSD17B7–UCK2
PBX1–B4GALT3
PBX1–NDUFS2
PBX1–PIGM
PBX1–PPOX
PBX1–SDHC
PBX1–UAP1
PBX1–UCK2
PLEKHO1–ACP6
SDHC–MGST3
SDHC–UCK2
OV
UCEC
LUSC
CESC
HNSC
CORD
PRAD
GBM
STAD
LGG
SKCM
BLCA
LUAD
LIHC
BRCA

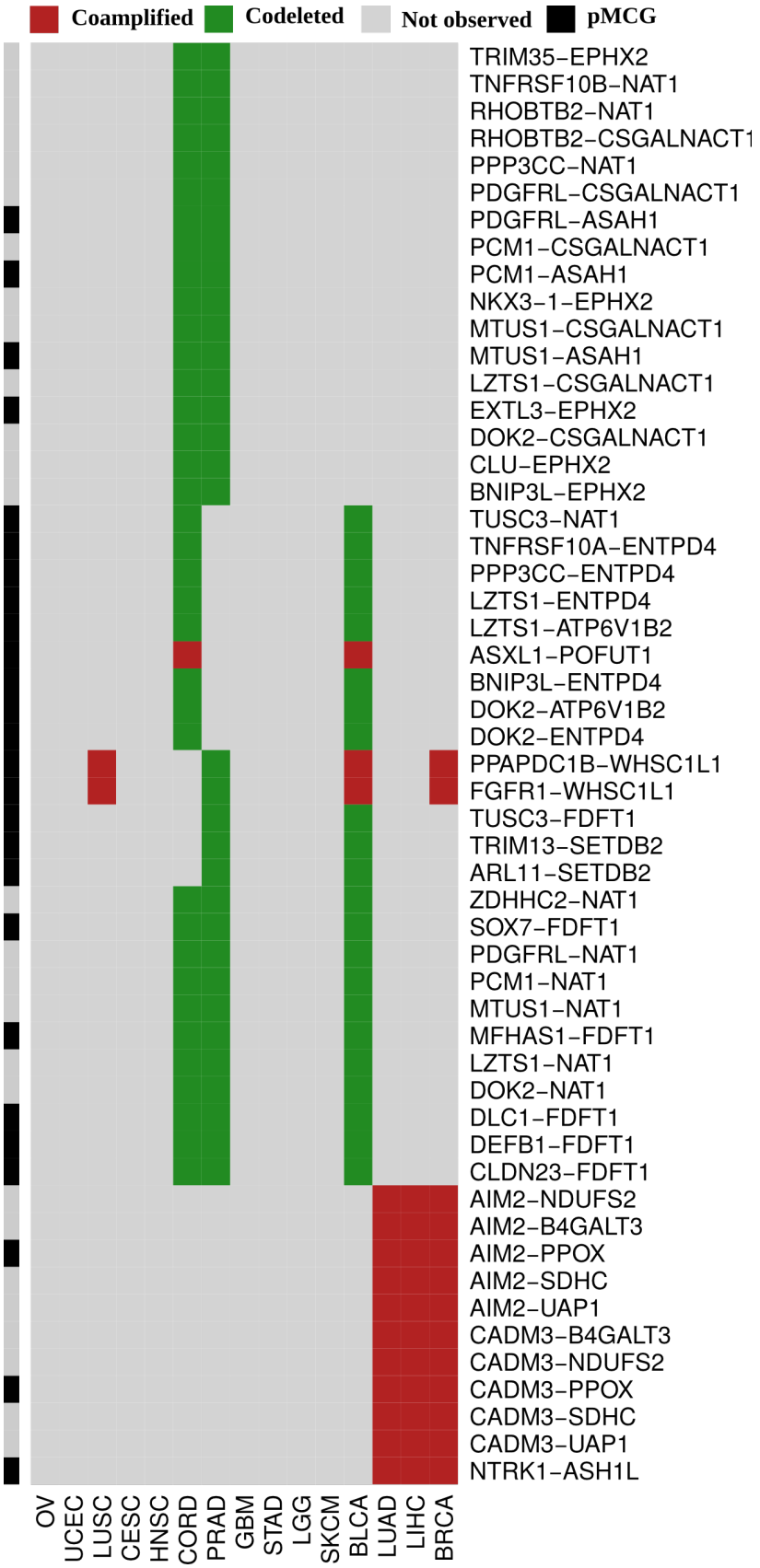
....continued


Coamplified
Codeleted
Not observed
pMCG
IL7R–AMACR
DAB2–OXCT1
IL7R–TARS
LIFR–AMACR
LIFR–OXCT1
LIFR–HMGCS1
DAB2–NNT
SFRP1–AGPAT6
AHRR–NDUFS6
AHRR–LPCAT1
SDHA–LPCAT1
AHRR–SDHA
SDHA–NDUFS6
DAB2–HMGCS1
CDKN2B–MTAP
CDKN2A–MTAP
IRX1–NDUFS6
IRX1–LPCAT1
IRX1–SDHA
IRX1–SRD5A1
PDE4DIP–PI4KB
BCL9–PI4KB
PLEKHO1–CA14
PDE4DIP–CA14
BCL9–CA14
S100A11–CA14
TXNIP–CA14
MUC1–SETDB1
MUC1–PRUNE
RIT1–PI4KB
TPM3–PIP5K1A
TPM3–PRUNE
EIF4A2–ST6GAL1
BCL6–ST6GAL1
ETV5–ST6GAL1
LPP–ST6GAL1
SERPINI2–PLD1
TP63–ST6GAL1
FBXO32–NDUFB9
ATAD2–NDUFB9
MTSS1–NDUFB9
MYC–NDUFB9
MTSS1–SQLE
ATAD2–SQLE
FBXO32–SQLE
MYC–SQLE
PLEKHA6–ETNK2
BTG2–ETNK2
MDM4–ETNK2
KISS1–ETNK2
PPP1R1B–PNMT
ERBB2–PNMT
RECQL4–DGAT1
OV
UCEC
LUSC
CESC
HNSC
CORD
PRAD
GBM
STAD
LGG
SKCM
BLCA
LUAD
LIHC
BRCA

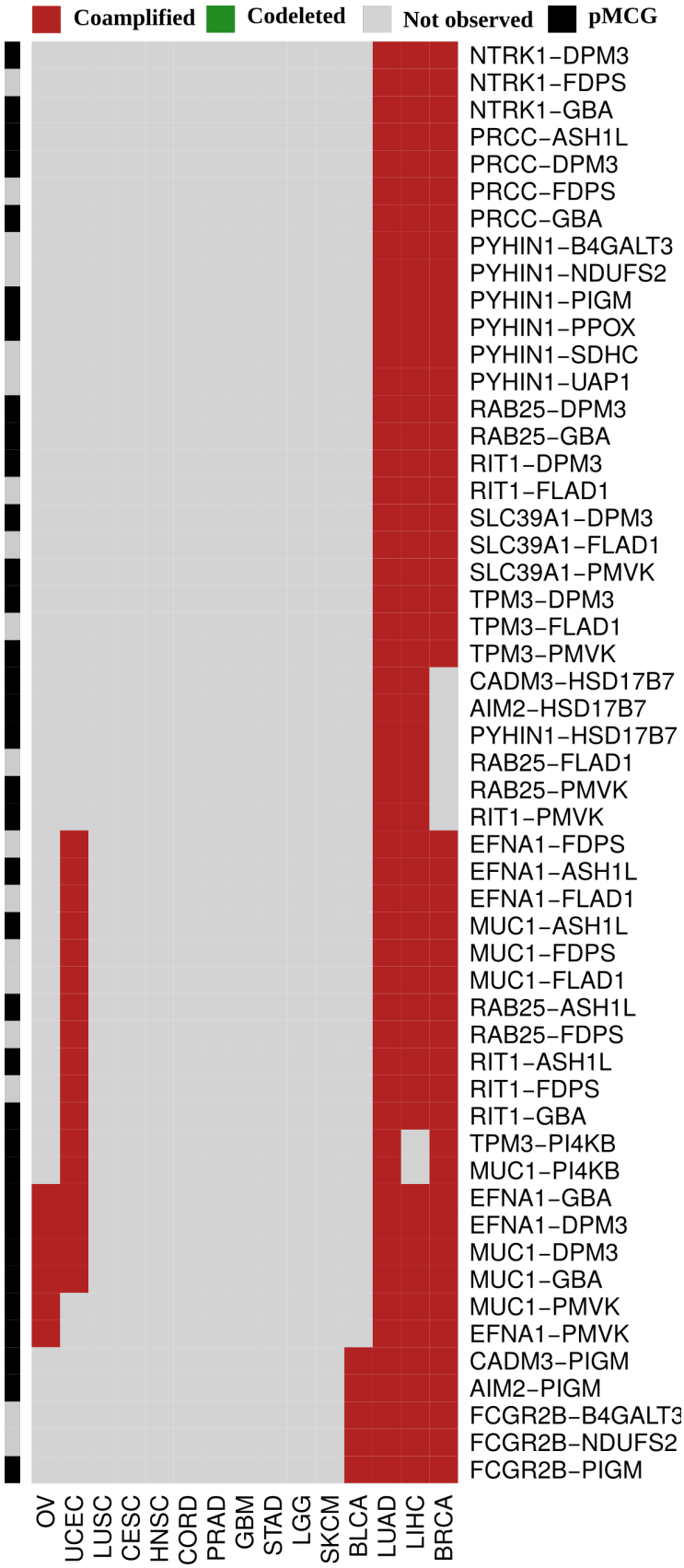
....continued


Coamplified
Codeleted
Not observed
pMCG
RECQL4–CYC1
RECQL4–GPAA1
RECQL4–TSTA3
SCRIB–CYC1
SCRIB–DGAT1
SCRIB–GPAA1
SCRIB–TSTA3
NDRG1–ST3GAL1
SCRIB–OPLAH
RECQL4–OPLAH
SCRIB–PYCRL
RECQL4–PYCRL
RAD21–EXT1
SIRT2–SARS2
MAP4K1–SARS2
SPINT2–SARS2
PAF1–SARS2
ATR–CHST2
ZFP36–SARS2
PDE4DIP–ACP6
BCL9–ACP6
TXNIP–ACP6
CCNE1–UQCRFS1
GTPBP4–IDI1
TNFRSF10B–ENTPD4
RHOBTB2–ENTPD4
NKX3–1–ENTPD4
GADD45GIP1–GCDH
GADD45GIP1–FARSA
CCND2–NDUFA9
CCND2–C12orf5
EEF1E1–FARS2
BRAF–NDUFB2
FOXC1–FARS2
SCRIB–GPT
RECQL4–GPT
EGLN1–GALNT2
ARF1–ITPKB
EGLN1–GGPS1
EGLN1–GNPAT
FH–ADSS
H3F3A–EPHX1
H3F3A–ITPKB
DIDO1–UCKL1
CDH4–UCKL1
SS18L1–UCKL1
GNAS–ATP5E
CABLES2–UCKL1
SDC4–ACOT8
SDC4–PIGT
ZDHHC2–CSGALNACT1
ZDHHC2–ASAH1
TUSC3–CSGALNACT1
OV
UCEC
LUSC
CESC
HNSC
CORD
PRAD
GBM
STAD
LGG
SKCM
BLCA
LUAD
LIHC
BRCA

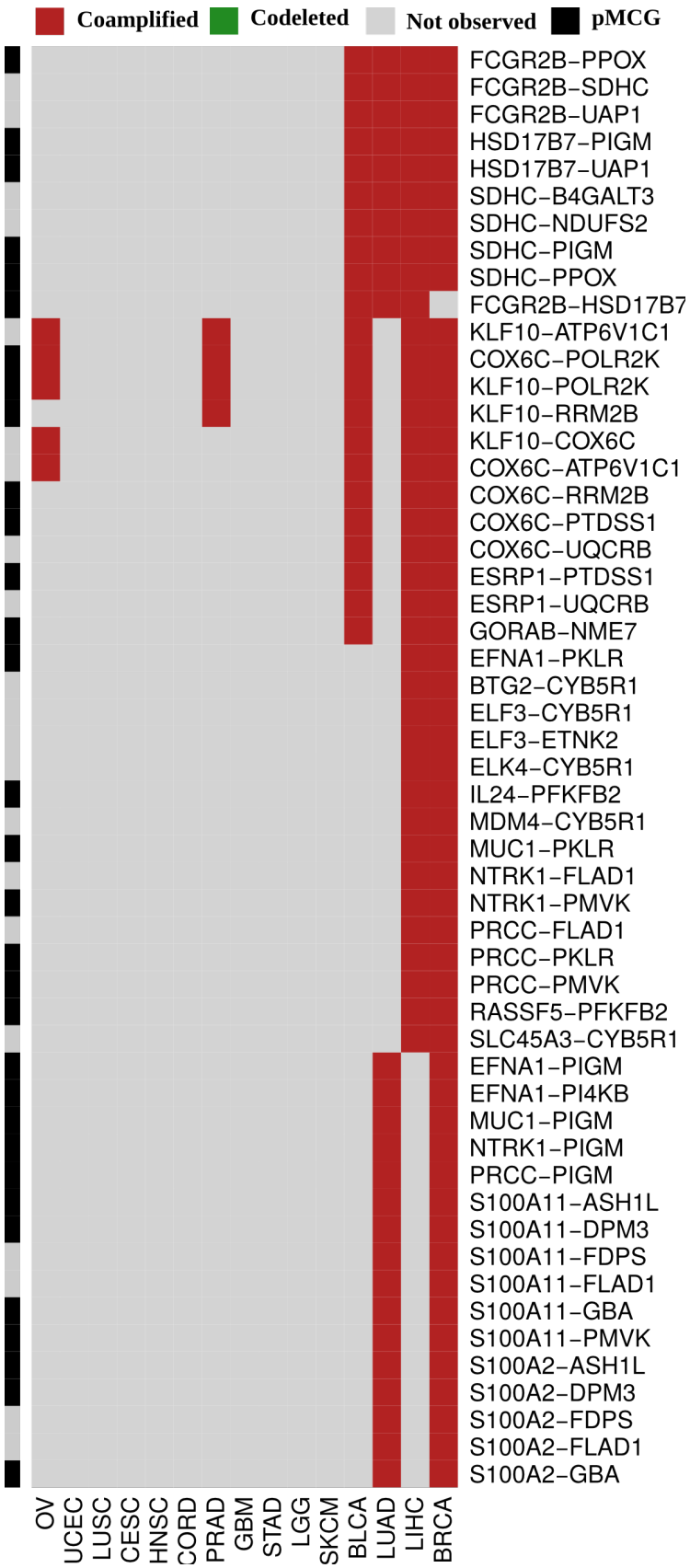
....continued


Coamplified
Codeleted
Not observed
pMCG
TRIM35–EPHX2
TNFRSF10B–NAT1
RHOBTB2–NAT1
RHOBTB2–CSGALNACT1
PPP3CC–NAT1
PDGFRL–CSGALNACT1
PDGFRL–ASAH1
PCM1–CSGALNACT1
PCM1–ASAH1
NKX3–1–EPHX2
MTUS1–CSGALNACT1
MTUS1–ASAH1
LZTS1–CSGALNACT1
EXTL3–EPHX2
DOK2–CSGALNACT1
CLU–EPHX2
BNIP3L–EPHX2
TUSC3–NAT1
TNFRSF10A–ENTPD4
PPP3CC–ENTPD4
LZTS1–ENTPD4
LZTS1–ATP6V1B2
ASXL1–POFUT1
BNIP3L–ENTPD4
DOK2–ATP6V1B2
DOK2–ENTPD4
PPAPDC1B–WHSC1L1
FGFR1–WHSC1L1
TUSC3–FDFT1
TRIM13–SETDB2
ARL11–SETDB2
ZDHHC2–NAT1
SOX7–FDFT1
PDGFRL–NAT1
PCM1–NAT1
MTUS1–NAT1
MFHAS1–FDFT1
LZTS1–NAT1
DOK2–NAT1
DLC1–FDFT1
DEFB1–FDFT1
CLDN23–FDFT1
AIM2–NDUFS2
AIM2–B4GALT3
AIM2–PPOX
AIM2–SDHC
AIM2–UAP1
CADM3–B4GALT3
CADM3–NDUFS2
CADM3–PPOX
CADM3–SDHC
CADM3–UAP1
NTRK1–ASH1L
OV
UCEC
LUSC
CESC
HNSC
CORD
PRAD
GBM
STAD
LGG
SKCM
BLCA
LUAD
LIHC
BRCA

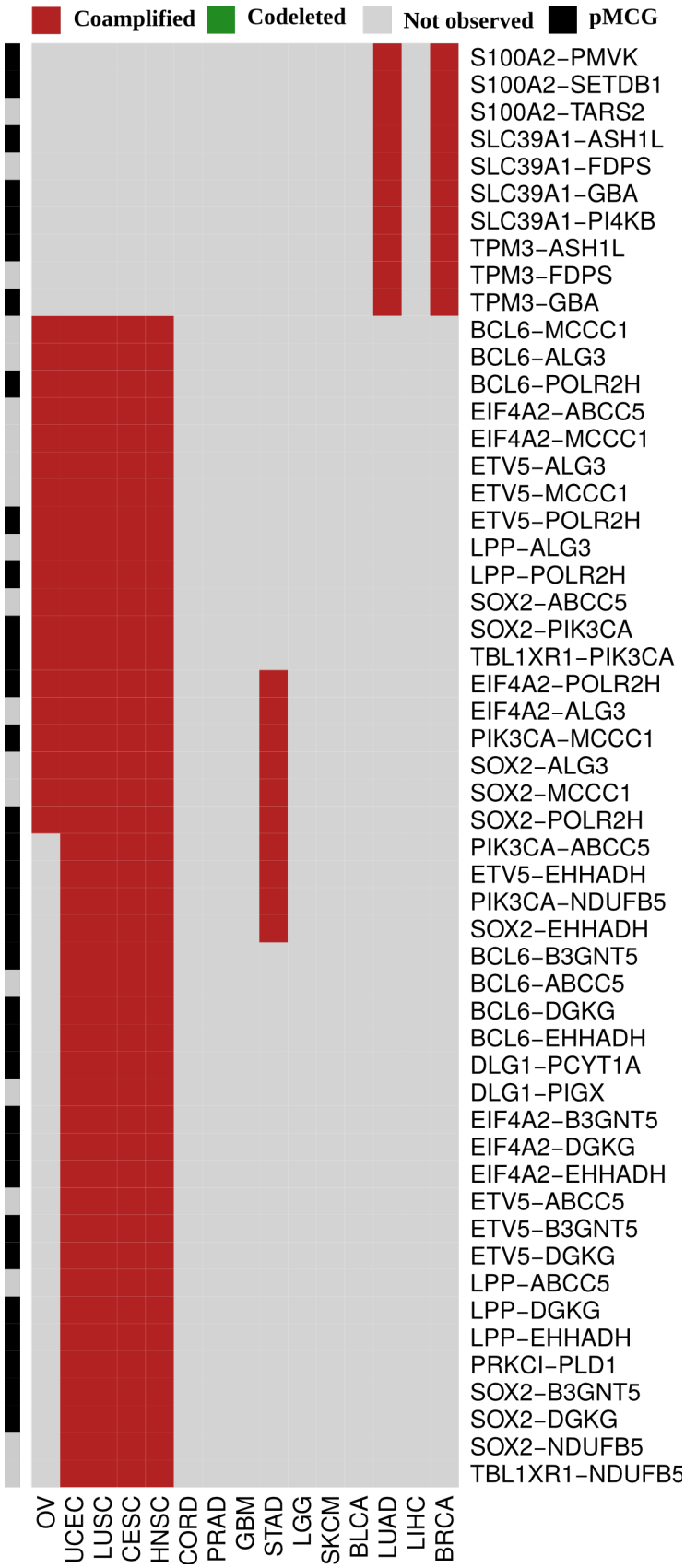
....continued


Coamplified
Codeleted
Not observed
pMCG
NTRK1–DPM3
NTRK1–FDPS
NTRK1–GBA
PRCC–ASH1L
PRCC–DPM3
PRCC–FDPS
PRCC–GBA
PYHIN1–B4GALT3
PYHIN1–NDUFS2
PYHIN1–PIGM
PYHIN1–PPOX
PYHIN1–SDHC
PYHIN1–UAP1
RAB25–DPM3
RAB25–GBA
RIT1–DPM3
RIT1–FLAD1
SLC39A1–DPM3
SLC39A1–FLAD1
SLC39A1–PMVK
TPM3–DPM3
TPM3–FLAD1
TPM3–PMVK
CADM3–HSD17B7
AIM2–HSD17B7
PYHIN1–HSD17B7
RAB25–FLAD1
RAB25–PMVK
RIT1–PMVK
EFNA1–FDPS
EFNA1–ASH1L
EFNA1–FLAD1
MUC1–ASH1L
MUC1–FDPS
MUC1–FLAD1
RAB25–ASH1L
RAB25–FDPS
RIT1–ASH1L
RIT1–FDPS
RIT1–GBA
TPM3–PI4KB
MUC1–PI4KB
EFNA1–GBA
EFNA1–DPM3
MUC1–DPM3
MUC1–GBA
MUC1–PMVK
EFNA1–PMVK
CADM3–PIGM
AIM2–PIGM
FCGR2B–B4GALT3
FCGR2B–NDUFS2
FCGR2B–PIGM
OV
UCEC
LUSC
CESC
HNSC
CORD
PRAD
GBM
STAD
LGG
SKCM
BLCA
LUAD
LIHC
BRCA

....continued


Coamplified
Codeleted
Not observed
pMCG
FCGR2B–PPOX
FCGR2B–SDHC
FCGR2B–UAP1
HSD17B7–PIGM
HSD17B7–UAP1
SDHC–B4GALT3
SDHC–NDUFS2
SDHC–PIGM
SDHC–PPOX
FCGR2B–HSD17B7
KLF10–ATP6V1C1
COX6C–POLR2K
KLF10–POLR2K
KLF10–RRM2B
KLF10–COX6C
COX6C–ATP6V1C1
COX6C–RRM2B
COX6C–PTDSS1
COX6C–UQCRB
ESRP1–PTDSS1
ESRP1–UQCRB
GORAB–NME7
EFNA1–PKLR
BTG2–CYB5R1
ELF3–CYB5R1
ELF3–ETNK2
ELK4–CYB5R1
IL24–PFKFB2
MDM4–CYB5R1
MUC1–PKLR
NTRK1–FLAD1
NTRK1–PMVK
PRCC–FLAD1
PRCC–PKLR
PRCC–PMVK
RASSF5–PFKFB2
SLC45A3–CYB5R1
EFNA1–PIGM
EFNA1–PI4KB
MUC1–PIGM
NTRK1–PIGM
PRCC–PIGM
S100A11–ASH1L
S100A11–DPM3
S100A11–FDPS
S100A11–FLAD1
S100A11–GBA
S100A11–PMVK
S100A2–ASH1L
S100A2–DPM3
S100A2–FDPS
S100A2–FLAD1
S100A2–GBA
OV
UCEC
LUSC
CESC
HNSC
CORD
PRAD
GBM
STAD
LGG
SKCM
BLCA
LUAD
LIHC
BRCA

....continued


Coamplified
Codeleted
Not observed
pMCG
S100A2–PMVK
S100A2–SETDB1
S100A2–TARS2
SLC39A1–ASH1L
SLC39A1–FDPS
SLC39A1–GBA
SLC39A1–PI4KB
TPM3–ASH1L
TPM3–FDPS
TPM3–GBA
BCL6–MCCC1
BCL6–ALG3
BCL6–POLR2H
EIF4A2–ABCC5
EIF4A2–MCCC1
ETV5–ALG3
ETV5–MCCC1
ETV5–POLR2H
LPP–ALG3
LPP–POLR2H
SOX2–ABCC5
SOX2–PIK3CA
TBL1XR1–PIK3CA
EIF4A2–POLR2H
EIF4A2–ALG3
PIK3CA–MCCC1
SOX2–ALG3
SOX2–MCCC1
SOX2–POLR2H
PIK3CA–ABCC5
ETV5–EHHADH
PIK3CA–NDUFB5
SOX2–EHHADH
BCL6–B3GNT5
BCL6–ABCC5
BCL6–DGKG
BCL6–EHHADH
DLG1–PCYT1A
DLG1–PIGX
EIF4A2–B3GNT5
EIF4A2–DGKG
EIF4A2–EHHADH
ETV5–ABCC5
ETV5–B3GNT5
ETV5–DGKG
LPP–ABCC5
LPP–DGKG
LPP–EHHADH
PRKCI–PLD1
SOX2–B3GNT5
SOX2–DGKG
SOX2–NDUFB5
TBL1XR1–NDUFB5
OV
UCEC
LUSC
CESC
HNSC
CORD
PRAD
GBM
STAD
LGG
SKCM
BLCA
LUAD
LIHC
BRCA

....continued

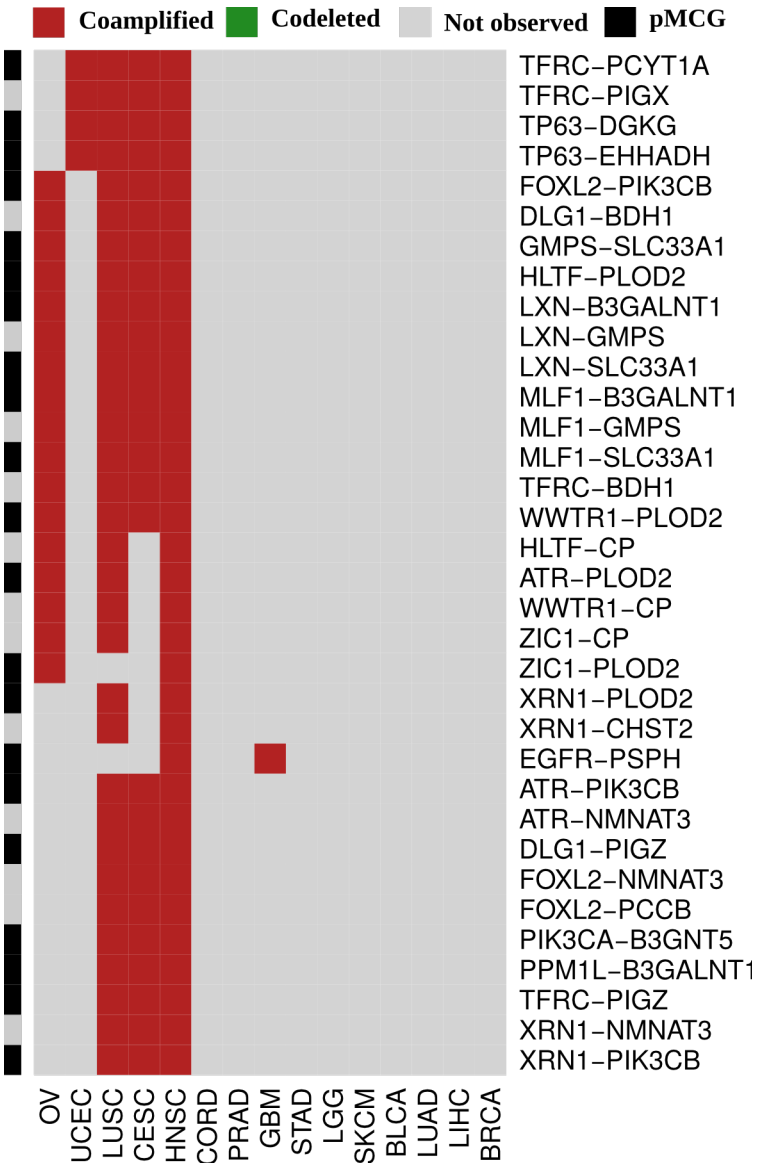
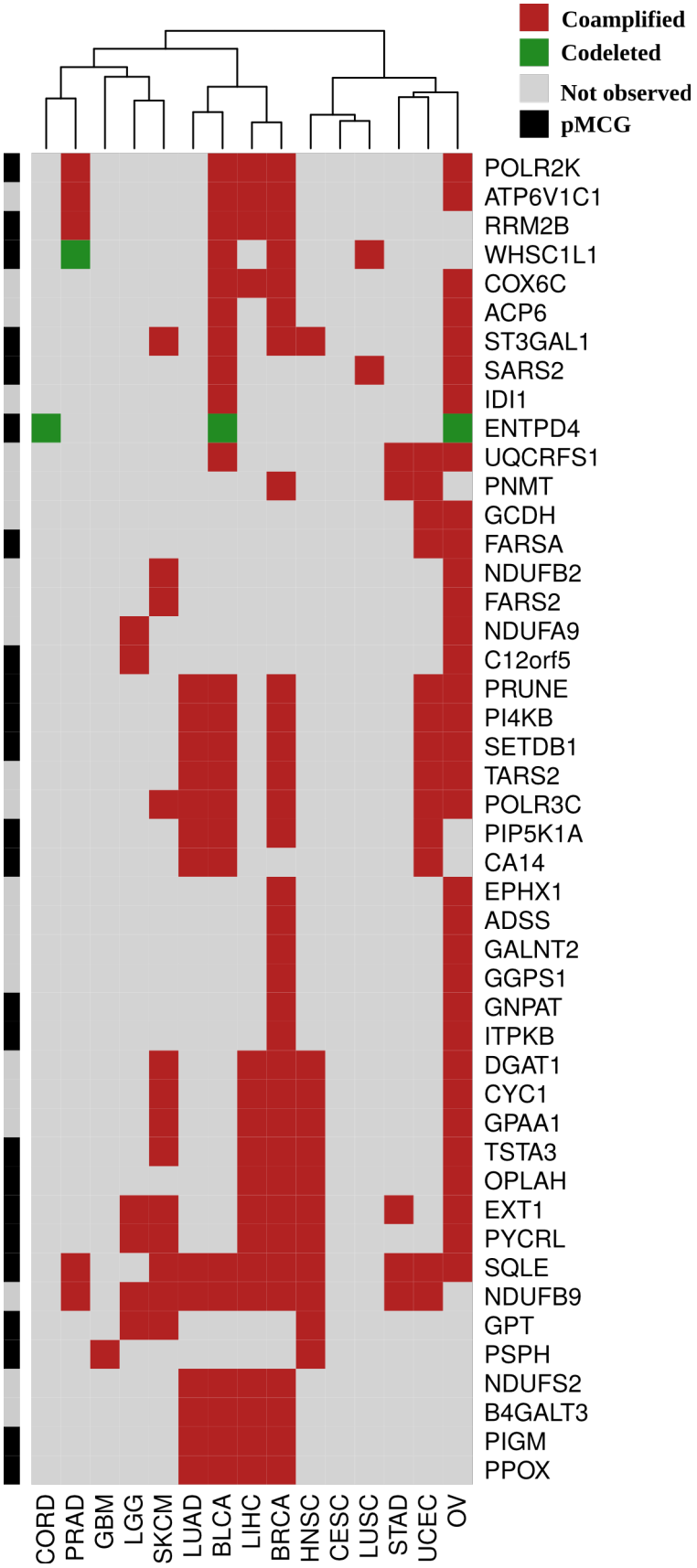
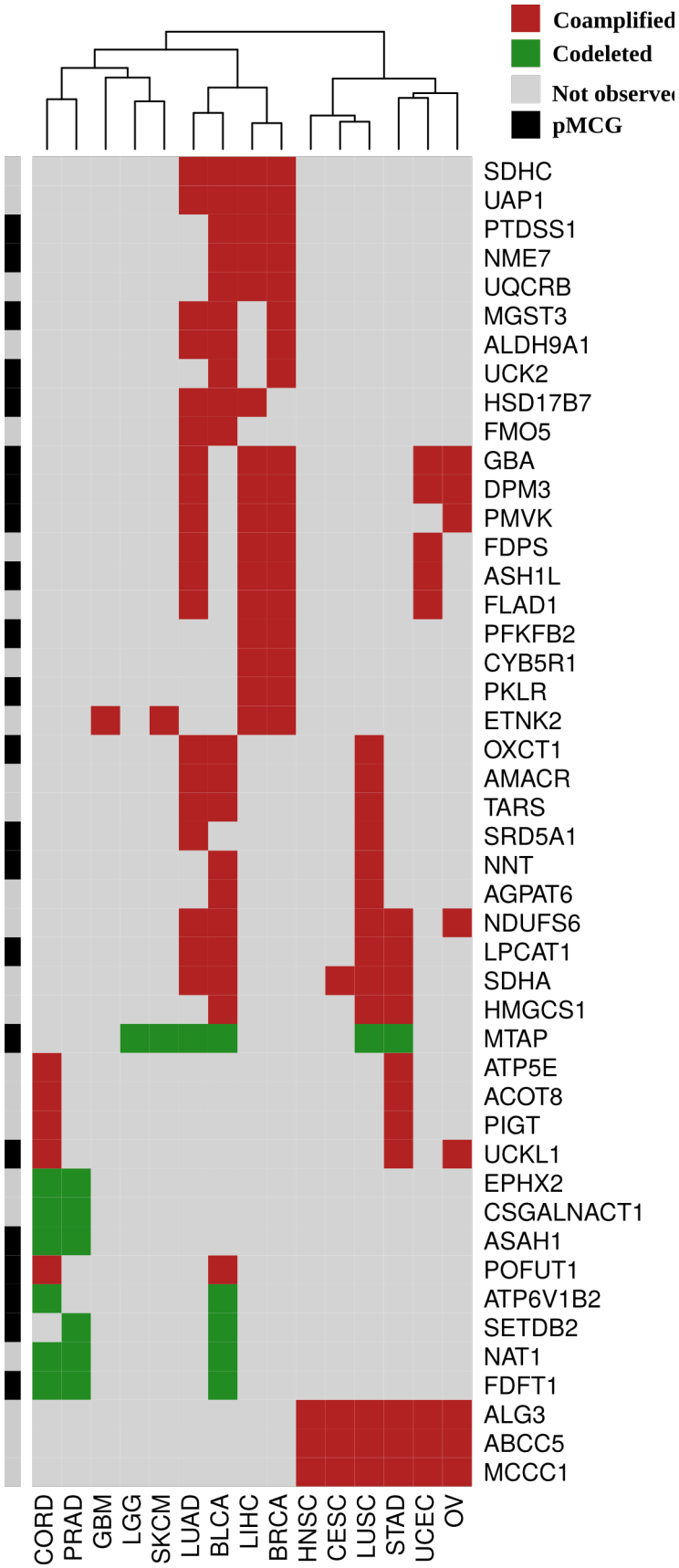


Figure 30: **Heatmap of all statistically prioritized and expression integrated cancer-metabolic gene pairs.** Heatmap of all cancer-metabolic gene pairs (N=405) identified after statistical priorization and co-expression analysis in at least 2 cancer types using the *iMetCG* pipeline. This heatmap shows the complete data of Fig. 19. Red and green boxes represent amplification and deletion events, respectively. The column bar on the left highlights the putative metabolic cancer genes (in black) identified in later functional prioritization steps.


Coamplified
Codeleted
Not observed
pMCG
POLR2K
ATP6V1C1
RRM2B
WHSC1L1
COX6C
ACP6
ST3GAL1
SARS2
IDI1
ENTPD4
UQCRFS1
PNMT
GCDH
FARSA
NDUFB2
FARS2
NDUFA9
C12orf5
PRUNE
PI4KB
SETDB1
TARS2
POLR3C
PIP5K1A
CA14
EPHX1
ADSS
GALNT2
GGPS1
GNPAT
ITPKB
DGAT1
CYC1
GPAA1
TSTA3
OPLAH
EXT1
PYCRL
SQLE
NDUFB9
GPT
PSPH
NDUFS2
B4GALT3
PIGM
PPOX
CORD
PRAD
GBM
LGG
SKCM
LUAD
BLCA
LIHC
BRCA
HNSC
CESC
LUSC
STAD
UCEC
OV

....continued


Coamplified
Codeleted
Not observe
pMCG
SDHC
UAP1
PTDSS1
NME7
UQCRB
MGST3
ALDH9A1
UCK2
HSD17B7
FMO5
GBA
DPM3
PMVK
FDPS
ASH1L
FLAD1
PFKFB2
CYB5R1
PKLR
ETNK2
OXCT1
AMACR
TARS
SRD5A1
NNT
AGPAT6
NDUFS6
LPCAT1
SDHA
HMGCS1
MTAP
ATP5E
ACOT8
PIGT
UCKL1
EPHX2
CSGALNACT1
ASAH1
POFUT1
ATP6V1B2
SETDB2
NAT1
FDFT1
ALG3
ABCC5
MCCC1
CORD
PRAD
GBM
LGG
SKCM
LUAD
BLCA
LIHC
BRCA
HNSC
CESC
LUSC
STAD
UCEC
OV

....continued

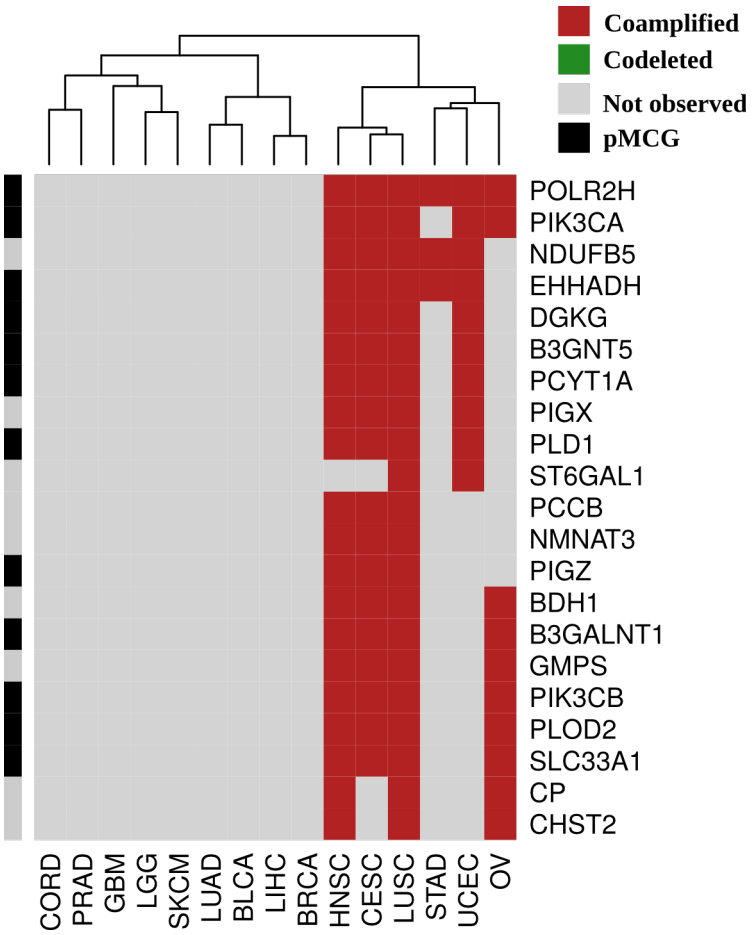


Figure 31: **Heatmap of the core metabolic gene set.** The identified metabolic genes found to be significantly co-altered in at least two cancer types (N=113) irrespective of its co-altered cancer gene partners. This heatmap shows the complete data of Fig. 22. The underlying data was used for clustering of the tumors (Fig. 23). Red and green boxes represent amplification and deletion events, respectively. The bar on the left shows the putative metabolic cancer genes (in black) identified in later steps.

BIBLIOGRAPHY

Cancer research uk, 2015. http://www.cancerresearchuk.org/health-professional/cancer-statistics/risk. Accessed: 2015-09-15.

Christopher S Ahn and Christian M Metallo. Mitochondria as biosynthetic factories for cancer proliferation. *Cancer & metabolism*, 3(1):1, 2015. (Cited on page 16.)

Fátima Al-Shahrour, Pablo Minguez, Tomás Marqués-Bonet, Elodie Gazave, Arcadi Navarro, et al. Selection upon genome architecture: conservation of functional neighborhoods with changing genes. *PLoS Comput Biol*, 6(10):e1000953, 2010. (Cited on page 27.)

Marco Archetti. Evolutionary dynamics of the warburg effect: glycolysis as a collective action problem among cancer cells. *Journal of theoretical biology*, 341:1–8, 2014. (Cited on page 18.)

K. Athenstaedt and G. Daum. Phosphatidic acid, a key intermediate in lipid metabolism. *Eur J Biochem*, 266(1):1–16, 1999. (Cited on pages 16 and 59.)

A. Bashashati, G. Haffari, J. Ding, G. Ha, K. Lui, J. Rosner, D. G. Huntsman, C. Caldas, S. A. Aparicio, and S. P. Shah. Drivernet: uncovering the impact of somatic driver mutations on transcriptional networks in cancer. *Genome Biol*, 13(12):R124, 2012. (Cited on pages 2 and 30.)

Nizar N Batada, Araxi O Urrutia, and Laurence D Hurst. Chromatin remodelling is a major source of coexpression of linked genes in yeast. *Trends in Genetics*, 23(10):480–484, 2007. (Cited on page 26.)

Yoav Benjamini and Yosef Hochberg. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 289–300, 1995. (Cited on page 40.)

Karim Bensaad, Atsushi Tsuruta, Mary A Selak, M Nieves Calvo Vidal, Katsunori Nakano, Ramon Bartrons, Eyal Gottlieb, and Karen H Vousden. Tigar, a p53-inducible regulator of glycolysis and apoptosis. *Cell*, 126(1):107–120, 2006. (Cited on page 23.)

J.M. Berg, J.L. Tymoczko, and L. Stryer. *Biochemistry, Fifth Edition*. W.H. Freeman, 2002. ISBN 9780716730514. (Cited on pages 10, 11, and 14.)

J. J. Berman. Tumor classification: molecular analysis meets aristotle. *BMC Cancer*, 4:10, 2004. (Cited on page 55.)

Joseph R. Bertino, William R. Waud, William B. Parker, and Martin Lubin. Targeting tumors that lack methylthioadenosine phosphorylase (mtap) activity. *Cancer Biology 'I&' Therapy*, 11(7):627–632, 2014. (Cited on pages 2, 30, and 59.)

Daniel C Berwick, Ingeborg Hers, Kate J Heesom, S Kelly Moule, and Jeremy M Tavaré. The identification of atp-citrate lyase as a protein kinase b (akt) substrate in primary adipocytes. *Journal of Biological Chemistry*, 277(37):33895–33900, 2002. (Cited on page 22.)

A. Bezginov, G. W. Clark, R. L. Charlebois, V. U. Dar, and E. R. Tillier. Coevolution reveals a network of human proteins originating with multicellularity. *Mol. Biol. Evol.*, 30(2):332–346, Feb 2013. (Cited on page 64.)

Stefan Bordag. A comparison of co-occurrence and similarity measures as simulations of context. In *Proceedings of the 9th International Conference on Computational Linguistics and Intelligent Text Processing*, CICLing'08, pages 52–63, Berlin, Heidelberg, 2008. Springer-Verlag. ISBN 3-540-78134-X, 978-3-540-78134-9. (Cited on page 39.)

G. Bouma. Normalized (pointwise) mutual information in collocation extraction. In *From Form to Meaning: Processing Texts Automatically, Proceedings of the Biennial GSCL Conference 2009*, pages 31–40, Tübingen, 2009. (Cited on page 38.)

Theodor Boveri. Concerning the origin of malignant tumours by theodor boveri. translated and annotated by henry harris. *Journal of Cell Science*, 121(Supplement 1):1–84, 2008. (Cited on page 5.)

Rob A Cairns, Isaac S Harris, and Tak W Mak. Regulation of cancer cell metabolism. *Nature Reviews Cancer*, 11(2):85–95, 2011. (Cited on pages 6, 12, 13, 22, and 25.)

Isabela Cajiao, Aiwen Zhang, Eung Jae Yoo, Nancy E Cooke, and Stephen A Liebhaber. Bystander gene activation by a locus control region. *The EMBO journal*, 23(19):3854–3863, 2004. (Cited on page 27.)

H. Caron, B. van Schaik, M. van der Mee, F. Baas, G. Riggins, P. van Sluis, M. C. Hermus, R. van Asperen, K. Boon, P. A. Voute, S. Heisterkamp, A. van Kampen, and R. Versteeg. The human transcriptome map: clustering of highly expressed genes in chromosomal domains. *Science*, 291(5507):1289–92, 2001. (Cited on page 2.)

D. A. Carson, T. Nobori, E. O. Kajander, C. J. Carrera, M. Kubota, and H. Yamanaka. Methylthioadenosine (mesado) phosphorylase defi-

ciency in malignancy. *Adv Exp Med Biol*, 250:179–85, 1988. (Cited on pages 2 and 30.)

Olga Catalina-Rodriguez, Vamsi Kolukula, York Tomita, Anju Preet, Ferdinando Palmieri, Anton Wellstein, Stephen Byers, Amato Giaccia, Eric Glasgow, Chris Albanese, and Maria Laura Avantaggiati. The mitochondrial citrate transporter, cic, is essential for mitochondrial homeostasis. *Oncotarget*, 3(10), 2012. (Cited on page 19.)

Barbara Chaneton, Petra Hillmann, Liang Zheng, Agnès CL Martin, Oliver DK Maddocks, Achuthanunni Chokkathukalam, Joseph E Coyle, Andris Jankevics, Finn P Holding, Karen H Vousden, et al. Serine is a natural ligand and allosteric activator of pyruvate kinase m2. *Nature*, 491(7424):458–462, 2012. (Cited on page 25.)

Lilan Chen, Ting Liu, Jinhua Zhou, Yunfei Wang, Xinran Wang, Wen Di, and Shu Zhang. Citrate synthase expression affects tumor phenotype and drug resistance in human ovarian carcinoma. *PLoS ONE*, 9(12):e115708, 12 2014. (Cited on page 19.)

W. H. Chen, P. Minguez, M. J. Lercher, and P. Bork. Ogee: an online gene essentiality database. *Nucleic Acids Res*, 40(Database issue): D901–6, 2012. (Cited on page 35.)

F. X. Cheng, P. L. Jia, Q. Wang, C. C. Lin, W. H. Li, and Z. M. Zhao. Studying tumorigenesis through network evolution and somatic mutational perturbations in the cancer interactome. *Molecular Biology and Evolution*, 31(8):2156–2169, 2014. (Cited on page 63.)

Adam L Cohen, Sheri L Holmen, and Howard Colman. Idh1 and idh2 mutations in gliomas. *Current neurology and neuroscience reports*, 13 (5):1–7, 2013. (Cited on page 25.)

B. A. Cohen, R. D. Mitra, J. D. Hughes, and G. M. Church. A computational analysis of whole-genome expression data reveals chromosomal domains of gene expression. *Nat Genet*, 26(2):183–6, 2000. (Cited on page 2.)

Marcus S. Cooke, Mark D. Evans, Miral Dizdaroglu, and Joseph Lunec. Oxidative dna damage: mechanisms, mutation, and disease. *The FASEB Journal*, 17(10):1195–1214, 2003. (Cited on page 13.)

Athel Cornish-Bowden. Current iubmb recommendations on enzyme nomenclature and kinetics. *Perspectives in Science*, 1(1):74–87, 2014. (Cited on page 8.)

John T Cunningham, Melissa V Moreno, Alessia Lodi, Sabrina M Ronen, and Davide Ruggero. Protein and nucleotide biosynthesis are coupled by a single rate-limiting enzyme, prps2, to drive cancer. *Cell*, 157(5):1088–1103, 2014. (Cited on page 23.)

Erin Currie, Almut Schulze, Rudolf Zechner, Tobias C. Walther, and Robert V. Farese. Cellular fatty acid metabolism and cancer. 18(2): 153–161, 2013. (Cited on pages 19 and 21.)

Chi V Dang. Links between metabolism and cancer. *Genes & development*, 26(9):877–890, 2012. (Cited on pages 12 and 17.)

M. Davila Lopez, J. J. Martinez Guerra, and T. Samuelsson. Analysis of gene order conservation in eukaryotes identifies transcriptionally and functionally linked genes. *PLoS One*, 5(5):e10654, 2010. (Cited on pages 2 and 27.)

T. Davoli, A. W. Xu, K. E. Mengwasser, L. M. Sack, J. C. Yoon, P. J. Park, and S. J. Elledge. Cumulative haploinsufficiency and triplosensitivity drive aneuploidy patterns and shape the cancer genome. *Cell*, 155(4):948–62, 2013. (Cited on pages 2, 26, 28, 29, 35, and 65.)

Ralph J DeBerardinis and Tzuling Cheng. Q's next: the diverse functions of glutamine in metabolism, cell biology and cancer. *Oncogene*, 29(3):313–324, 2010. (Cited on page 18.)

Ralph J DeBerardinis, Anthony Mancuso, Evgueni Daikhin, Ilana Nissim, Marc Yudkoff, Suzanne Wehrli, and Craig B Thompson. Beyond aerobic glycolysis: transformed cells can engage in glutamine metabolism that exceeds the requirement for protein and nucleotide synthesis. *Proceedings of the National Academy of Sciences*, 104(49):19345–19350, 2007. (Cited on page 18.)

Ralph J DeBerardinis, Julian J Lum, Georgia Hatzivassiliou, and Craig B Thompson. The biology of cancer: metabolic reprogramming fuels cell growth and proliferation. *Cell metabolism*, 7(1):11–20, 2008. (Cited on pages 12 and 67.)

Johan Deprez, Didier Vertommen, Dario R Alessi, Louis Hue, and Mark H Rider. Phosphorylation and activation of heart 6-phosphofructo-2-kinase by protein kinase b and other protein kinases of the insulin signaling cascades. *Journal of Biological Chemistry*, 272(28):17269–17275, 1997. (Cited on page 22.)

Jill D Dombrauckas, Bernard D Santarsiero, and Andrew D Mesecar. Structural basis for tumor pyruvate kinase m2 allosteric regulation and catalysis. *Biochemistry*, 44(27):9417–9429, 2005. (Cited on page 25.)

Miki Ebisuya, Takuya Yamamoto, May Nakajima, and Eisuke Nishida. Ripples from neighbouring transcription. *Nature cell biology*, 10(9): 1106–1113, 2008. (Cited on page 26.)

Rebecca L Elstrom, Daniel E Bauer, Monica Buzzai, Robyn Karnauskas, Marian H Harris, David R Plas, Hongming Zhuang,

Ryan M Cinalli, Abass Alavi, Charles M Rudin, et al. Akt stimulates aerobic glycolysis in cancer cells. *Cancer research*, 64(11):3892–3899, 2004. (Cited on page 22.)

Jeffrey A Engelman, Ji Luo, and Lewis C Cantley. The evolution of phosphatidylinositol 3-kinases as regulators of growth and metabolism. *Nature Reviews Genetics*, 7(8):606–619, 2006. (Cited on page 22.)

Brandon Faubert, Emma E Vincent, Takla Griss, Bozena Samborska, Said Izreig, Robert U Svensson, Orval A Mamer, Daina Avizonis, David B Shackelford, Reuben J Shaw, et al. Loss of the tumor suppressor lkb1 promotes metabolic reprogramming of cancer cells via hif-1α. *Proceedings of the National Academy of Sciences*, 111(7):2554–2559, 2014. (Cited on page 24.)

Eric R. Fearon. Molecular genetics of colorectal cancer. *Annual Review of Pathology: Mechanisms of Disease*, 6(1):479–507, 2011. (Cited on page 28.)

Jacques Ferlay, Isabelle Soerjomataram, Rajesh Dikshit, Sultan Eser, Colin Mathers, Marise Rebelo, Donald Maxwell Parkin, David Forman, and Freddie Bray. Cancer incidence and mortality worldwide: Sources, methods and major patterns in globocan 2012. *International Journal of Cancer*, 136(5):E359–E386, 2015. (Cited on page 5.)

Richard Flavin, Stephane Peluso, Paul L Nguyen, and Massimo Loda. Fatty acid synthase as a potential therapeutic target in cancer. *Future oncology (London, England)*, 6(4):551–562, 2010. (Cited on page 21.)

D. A. Foster. Regulation of mtor by phosphatidic acid? *Cancer Res*, 67 (1):1–4, 2007. (Cited on pages 16 and 59.)

Lorenzo Galluzzi, Oliver Kepp, Matthew G Vander Heiden, and Guido Kroemer. Metabolic targets for cancer therapy. *Nature reviews Drug discovery*, 12(11):829–846, 2013. (Cited on pages 12 and 66.)

F. Gatto, I. Nookaew, and J. Nielsen. Chromosome 3p loss of heterozygosity is associated with a unique metabolic network in clear cell renal carcinoma. *Proceedings of the National Academy of Sciences of the United States of America*, 111(9):E866–E875, 2014. (Cited on pages 2 and 30.)

B. Georgi, B. F. Voight, and M. Bucan. From mouse to human: evolutionary genomics analysis of human orthologs of essential genes. *PLoS Genet*, 9(5):e1003484, 2013. (Cited on page 35.)

H. J. Gierman, M. H. Indemans, J. Koster, S. Goetze, J. Seppen, D. Geerts, R. van Driel, and R. Versteeg. Domain-wide regulation of

gene expression in the human genome. *Genome Res*, 17(9):1286–95, 2007. (Cited on page 2.)

G. V. Glinsky, A. Krones-Herzig, and A. B. Glinskii. Malignancy-associated regions of transcriptional activation: gene expression profiling identifies common chromosomal regions of a recurrent transcriptional activation in human prostate, breast, ovarian, and colon cancers. *Neoplasia*, 5(3):218–28, 2003. (Cited on pages 2 and 29.)

Gabriele V. Gnoni, Paola Priore, M. J H Geelen, and Luisa Siculella. The mitochondrial citrate carrier: Metabolic role and regulation of its activity and expression. 61(10):987–994, 2009. (Cited on page 19.)

Kathrin Gottlob, Nathan Majewski, Scott Kennedy, Eugene Kandel, R Brooks Robey, and Nissim Hay. Inhibition of early apoptotic events by akt/pkb is dependent on the first committed step of glycolysis and mitochondrial hexokinase. *Genes & development*, 15(11): 1406–1418, 2001. (Cited on page 22.)

Mel Greaves and Carlo C. Maley. Clonal evolution in cancer. *Nature*, 481(7381):306–313, 2012. (Cited on page 2.)

A.J.F. Griffiths, S.R. Wessler, S.B. Carroll, and J. Doebley. *Introduction to Genetic Analysis*. W. H. Freeman, 2010. ISBN 9781464100277. (Cited on page 6.)

Dan Y Gui, Caroline A Lewis, and Matthew G Vander Heiden. Allosteric regulation of pkm2 allows cellular adaptation to different physiological states. *Science signaling*, 6(263):pe7–pe7, 2013. (Cited on pages 21 and 25.)

Fabienne Guillaumond, Julie Leca, Orianne Olivares, Marie-Noëlle Lavaut, Nicolas Vidal, Patrice Berthezène, Nelson Javier Dusetti, Céline Loncle, Ezequiel Calvo, Olivier Turrini, et al. Strengthened glycolysis under hypoxia supports tumor symbiosis and hexosamine biosynthesis in pancreatic adenocarcinoma. *Proceedings of the National Academy of Sciences*, 110(10):3919–3924, 2013. (Cited on page 16.)

D. Hanahan and R. A. Weinberg. Hallmarks of cancer: the next generation. *Cell*, 144(5):646–74, 2011. (Cited on pages 1, 6, 7, and 13.)

Jun-ichi Hanai, Nathaniel Doro, Atsuo T Sasaki, Susumu Kobayashi, Lewis C Cantley, Pankaj Seth, and Vikas P Sukhatme. Inhibition of lung cancer growth: Atp citrate lyase knockdown and statin treatment leads to dual blockade of mitogen-activated protein kinase (mapk) and phosphatidylinositol-3-kinase (pi3k)/akt pathways. *Journal of cellular physiology*, 227(4):1709–1720, 2012. (Cited on page 20.)

T. Hart, K. R. Brown, F. Sircoulomb, R. Rottapel, and J. Moffat. Measuring error rates in genomic perturbation screens: gold standards for human functional genomics. *Mol Syst Biol*, 10:733, 2014. (Cited on page 35.)

Georgia Hatzivassiliou, Fangping Zhao, Daniel E. Bauer, Charalambos Andreadis, Anthony N. Shaw, Dashyant Dhanak, Sunil R. Hingorani, David A. Tuveson, and Craig B. Thompson. {ATP} citrate lyase inhibition can suppress tumor cell growth. *Cancer Cell*, 8(4): 311 – 321, 2005. (Cited on page 20.)

Matthew J Hayat, Nadia Howlader, Marsha E. Reichman, and Brenda K. Edwards. Cancer statistics, trends, and multiple primary cancer analyses from the surveillance, epidemiology, and end results (seer) program. *The oncologist*, 12(1):20–37, 2007. (Cited on page 5.)

Brian A Hemmings and David F Restuccia. Pi3k-pkb/akt pathway. *Cold Spring Harbor perspectives in biology*, 4(9):a011189, 2012. (Cited on page 22.)

Yugong Ho, Felice Elefant, Stephen A Liebhaber, and Nancy E Cooke. Locus control region transcription plays an active role in long-range gene activation. *Molecular cell*, 23(3):365–375, 2006. (Cited on page 27.)

K. A. Hoadley, C. Yau, D. M. Wolf, A. D. Cherniack, D. Tamborero, S. Ng, M. D. Leiserson, B. Niu, M. D. McLellan, V. Uzunangelov, J. Zhang, C. Kandoth, R. Akbani, H. Shen, L. Omberg, A. Chu, A. A. Margolin, L. J. Van't Veer, N. Lopez-Bigas, P. W. Laird, B. J. Raphael, L. Ding, A. G. Robertson, L. A. Byers, G. B. Mills, J. N. Weinstein, C. Van Waes, Z. Chen, E. A. Collisson, Network Cancer Genome Atlas Research, C. C. Benz, C. M. Perou, and J. M. Stuart. Multiplatform analysis of 12 cancer types reveals molecular classification within and across tissues of origin. *Cell*, 158(4):929–44, 2014. (Cited on pages 56 and 57.)

L. D. Hurst, C. Pal, and M. J. Lercher. The evolutionary dynamics of eukaryotic gene order. *Nat Rev Genet*, 5(4):299–310, 2004. (Cited on pages 2 and 26.)

H. Jeong, S. P. Mason, A. L. Barabasi, and Z. N. Oltvai. Lethality and centrality in protein networks. *Nature*, 411(6833):41–2, 2001. (Cited on page 62.)

Hawoong Jeong, Bálint Tombor, Réka Albert, Zoltan N Oltvai, and A-L Barabási. The large-scale organization of metabolic networks. *Nature*, 407(6804):651–654, 2000. (Cited on page 9.)

Peng Jiang, Wenjing Du, Anthony Mancuso, Kathryn E Wellen, and Xiaolu Yang. Reciprocal regulation of p53 and malic enzymes modulates metabolism and senescence. *Nature*, 493(7434):689–693, 2013. (Cited on page 16.)

Russell G Jones and Craig B Thompson. Tumor suppressors and cell metabolism: a recipe for cancer growth. *Genes & development*, 23(5): 537–548, 2009. (Cited on pages 21 and 23.)

C. Jose, N. Bellance, and R. Rossignol. Choosing between glycolysis and oxidative phosphorylation: a tumor's dilemma? *Biochim Biophys Acta*, 1807(6):552–61, 2011. (Cited on pages 17 and 59.)

Verline Justilien, Michael P Walsh, Syed A Ali, E Aubrey Thompson, Nicole R Murray, and Alan P Fields. The prkci and sox2 oncogenes are coamplified and cooperate to activate hedgehog signaling in lung squamous cell carcinoma. *Cancer cell*, 25(2):139–151, 2014. (Cited on page 30.)

Peter Kaatsch, Claudia Spix, Stefan Hentschel, Alexander Katalinic, Sabine Luttmann, and Christa Stegmaier. *Cancer in Germany 2009/2010(9th edition)*. Robert Koch Institute (ed.) and the Association of Population-based Cancer Registries in Germany (ed.), 2014. ISBN 978-3-89606-223-9. (Cited on page 5.)

M. Katoh. Epithelial-mesenchymal transition in gastric cancer (review). *Int J Oncol*, 27(6):1677–83, 2005. (Cited on page 57.)

Kelly M Kennedy and Mark W Dewhirst. Tumor metabolism of lactate: the influence and therapeutic potential for mct and cd147 regulation. *Future Oncology*, 6(1):127–148, 2010. (Cited on page 14.)

T. M. Kim, R. Xi, L. J. Luquette, R. W. Park, M. D. Johnson, and P. J. Park. Functional genomic analysis of chromosomal aberrations in a compendium of 8000 cancer genomes. *Genome Res*, 23(2):217–27, 2013. (Cited on pages 53, 56, and 66.)

M. Kimura. Preponderance of synonymous changes as evidence for the neutral theory of molecular evolution. *Nature*, 267(5608):275–6, 1977. (Cited on page 62.)

A King, MA Selak, , and E Gottlieb. Succinate dehydrogenase and fumarate hydratase: linking mitochondrial dysfunction and cancer. *Oncogene*, 25(34):4675–4682, 2006. (Cited on pages 21, 24, and 25.)

Christiaan Klijn, Jan Bot, David J Adams, Marcel Reinders, Lodewyk Wessels, and Jos Jonkers. Identification of networks of co-occurring, tumor-related dna copy number changes using a genome-wide scoring approach. *PLoS Comput Biol*, 6(1):e1000631–e1000631, 2010. (Cited on page 66.)

Alfred G Knudson. Mutation and cancer: statistical study of retinoblastoma. *Proceedings of the National Academy of Sciences*, 68 (4):820–823, 1971. (Cited on page 28.)

Rebecca Kohnken, Karthik M Kodigepalli, and Li Wu. Regulation of deoxynucleotide metabolism in cancer: novel mechanisms and therapeutic implications. *Molecular cancer*, 14(1):1, 2015. (Cited on page 25.)

WH Koppenol and PL Bounds. The warburg effect and metabolic efficiency: re-crunching the numbers. *Science*, 2009. (Cited on page 17.)

Willem H Koppenol, Patricia L Bounds, and Chi V Dang. Otto warburg's contributions to current concepts of cancer metabolism. *Nature Reviews Cancer*, 11(5):325–337, 2011. (Cited on pages 13 and 17.)

Preethi Korangath, Wei Wen Teo, Helen Sadik, Liangfeng Han, Noriko Mori, Charlotte M Huijts, Flonne Wildes, Santosh Bharti, Zhe Zhang, Cesar A Santa-Maria, et al. Targeting glutamine metabolism in breast cancer with aminooxyacetate. *Clinical Cancer Research*, pages clincanres–1200, 2015. (Cited on page 18.)

G. Kroemer and J. Pouyssegur. Tumor cell metabolism: cancer's achilles' heel. *Cancer Cell*, 13(6):472–82, 2008. (Cited on pages 3, 6, and 12.)

James R Krycer, Laura J Sharpe, Winnie Luu, and Andrew J Brown. The akt–srebp nexus: cell signaling meets lipid metabolism. *Trends in Endocrinology & Metabolism*, 21(5):268–276, 2010. (Cited on page 22.)

Lawrence N Kwong and Lynda Chin. Chromosome 10, frequently lost in human melanoma, encodes multiple tumor-suppressive functions. *Cancer research*, 74(6):1814–1821, 2014. (Cited on page 29.)

Anne Le, Charles R Cooper, Arvin M Gouw, Ramani Dinavahi, Anirban Maitra, Lorraine M Deck, Robert E Royer, David L Vander Jagt, Gregg L Semenza, and Chi V Dang. Inhibition of lactate dehydrogenase a induces oxidative stress and inhibits tumor progression. *Proceedings of the National Academy of Sciences*, 107(5):2037–2042, 2010. (Cited on page 14.)

J. M. Lee and E. L. Sonnhammer. Genomic gene clustering analysis of pathways in eukaryotes. *Genome Res*, 13(5):875–82, 2003. (Cited on pages 2 and 27.)

Marc Leibundgut, Timm Maier, Simon Jenni, and Nenad Ban. The multienzyme architecture of eukaryotic fatty acid synthases. 18(6): 714–725, 2008. (Cited on page 21.)

M. J. Lercher, A. O. Urrutia, and L. D. Hurst. Clustering of housekeeping genes provides a unified model of gene order in the human genome. *Nat Genet*, 31(2):180–3, 2002. (Cited on pages 2 and 27.)

Chin-Chih Lin, Tsung-Lin Cheng, Wen-Hui Tsai, Hui-Ju Tsai, Keng-Hsun Hu, Hao-Chun Chang, Chin-Wei Yeh, Ying-Chou Chen, Ching-Chun Liao, and Wen-Tsan Chang. Loss of the respiratory enzyme citrate synthase directly links the warburg effect to tumor malignancy. *Scientific reports*, 2, 2012. (Cited on page 19.)

Ruiting Lin, Xin Zhou, Wei Huang, Di Zhao, Lei Lv, Yue Xiong, Kun-Liang Guan, and Qun-Ying Lei. Acetylation control of cancer cell metabolism. *Current pharmaceutical design*, 20(15):2627–2633, 2014. (Cited on page 22.)

Yen-Chun Liu, Feng Li, Jesse Handler, CR Huang, Yan Xiang, Nicola Neretti, John M Sedivy, Karen I Zeller, and Chi V Dang. Global regulation of nucleotide biosynthetic genes by c-myc. *PloS one*, 3(7): e2722, 2008. (Cited on page 23.)

Jason W Locasale, Alexandra R Grassian, Tamar Melman, Costas A Lyssiotis, Katherine R Mattaini, Adam J Bass, Gregory Heffron, Christian M Metallo, Taru Muranen, Hadar Sharfi, et al. Phosphoglycerate dehydrogenase diverts glycolytic flux and contributes to oncogenesis. *Nature genetics*, 43(9):869–874, 2011. (Cited on page 16.)

Thomas M Loftus, Donna E Jaworsky, Gojeb L Frehywot, Craig A Townsend, Gabriele V Ronnett, M Daniel Lane, and Francis P Kuhajda. Reduced food intake and body weight in mice treated with fatty acid synthase inhibitors. *Science*, 288(5475):2379–2381, 2000. (Cited on page 21.)

H. Luo, Y. Lin, F. Gao, C. T. Zhang, and R. Zhang. Deg 10, an update of the database of essential genes that includes both protein-coding genes and noncoding genomic elements. *Nucleic Acids Res*, 42(Database issue):D574–80, 2014. (Cited on page 35.)

Ji Luo, Nicole L Solimini, and Stephen J Elledge. Principles of cancer therapy: oncogene and non-oncogene addiction. *Cell*, 136(5):823–837, 2009. (Cited on page 66.)

Ruth Lupu and Javier A Menendez. Pharmacological inhibitors of Fatty Acid Synthase (FASN)–catalyzed endogenous fatty acid biogenesis: a new family of anti-cancer agents? *Current pharmaceutical biotechnology*, 7(6):483–493, 2006. (Cited on page 21.)

Maria L Macheda, Suzanne Rogers, and James D Best. Molecular and cellular regulation of glucose transporter (glut) proteins in cancer. *Journal of cellular physiology*, 202(3):654–662, 2005. (Cited on page 14.)

Sudha Mannava, Vladimir Grachtchouk, Linda J Wheeler, Michael Im, Dazhong Zhuang, Elena G Slavina, Christopher K Mathews, Donna S Shewach, and Mikhail A Nikiforov. Direct role of nucleotide metabolism in c-myc-dependent proliferation of melanoma cells. *Cell Cycle*, 7(15):2392–2400, 2008. (Cited on page 23.)

Satoaki Matoba, Ju-Gyeong Kang, Willmar D Patino, Andrew Wragg, Manfred Boehm, Oksana Gavrilova, Paula J Hurley, Fred Bunz, and Paul M Hwang. p53 regulates mitochondrial respiration. *Science*, 312(5780):1650–1653, 2006. (Cited on page 24.)

Javier A Menendez and Ruth Lupu. Fatty acid synthase and the lipogenic phenotype in cancer pathogenesis. *Nature Reviews Cancer*, 7(10):763–777, 2007. (Cited on page 21.)

L. M. F. Merlo, J. W. Pepper, B. J. Reid, and C. C. Maley. Cancer as an evolutionary and ecological process. *Nature Reviews Cancer*, 6(12): 924–935, 2006. (Cited on page 2.)

Craig Mermel, Steven Schumacher, Barbara Hill, Matthew Meyerson, Rameen Beroukhim, and Gad Getz. Gistic2.0 facilitates sensitive and confident localization of the targets of focal somatic copy-number alteration in human cancers. *Genome Biology*, 12(4):R41, 2011. (Cited on page 33.)

Christian M Metallo and Matthew G Vander Heiden. Metabolism strikes back: metabolic flux regulates cell signaling. *Genes & development*, 24(24):2717–2722, 2010. (Cited on page 22.)

Christian M Metallo, Paulo A Gameiro, Eric L Bell, Katherine R Mattaini, Juanjuan Yang, Karsten Hiller, Christopher M Jewell, Zachary R Johnson, Darrell J Irvine, Leonard Guarente, et al. Reductive glutamine metabolism by idh1 mediates lipogenesis under hypoxia. *Nature*, 481(7381):380–384, 2012. (Cited on page 20.)

Rosalynn J Miltenberger, Kristine A Sukow, and Peggy J Farnham. An e-box-mediated increase in cad transcription at the g1/s-phase boundary is suppressed by inhibitory c-myc mutants. *Molecular and cellular biology*, 15(5):2527–2535, 1995. (Cited on page 23.)

I Mor, EC Cheung, and KH Vousden. Control of glycolysis through regulation of pfk1: old friends and recent additions. In *Cold Spring Harbor symposia on quantitative biology*, volume 76, pages 211–216. Cold Spring Harbor Laboratory Press, 2011. (Cited on page 14.)

S. Mukherjee. *The Emperor of All Maladies: A Biography of Cancer*. Scribner, 2010. ISBN 9781439181713. (Cited on page 5.)

Andrew R Mullen, William W Wheaton, Eunsook S Jin, Pei-Hsuan Chen, Lucas B Sullivan, Tzuling Cheng, Youfeng Yang, W Marston

Linehan, Navdeep S Chandel, and Ralph J DeBerardinis. Reductive carboxylation supports growth in tumour cells with defective mitochondria. *Nature*, 481(7381):385–388, 2012. (Cited on page 20.)

Maria E Mycielska, Ameet Patel, Nahit Rizaner, Maciej P Mazurek, Hector Keun, Anup Patel, Vadivel Ganapathy, and Mustafa B A Djamgoz. Citrate transport and metabolism in mammalian cells: prostate epithelial cells and prostate cancer. *BioEssays : news and reviews in molecular, cellular and developmental biology*, 31(1):10–20, 2009. (Cited on page 19.)

David L Nelson, Albert L Lehninger, and Michael M Cox. *Lehninger principles of biochemistry*. Macmillan, 2008. (Cited on pages 10, 11, 14, and 16.)

Cancer Genome Atlas Research Network, J. N. Weinstein, E. A. Collisson, G. B. Mills, K. R. Shaw, B. A. Ozenberger, K. Ellrott, I. Shmulevich, C. Sander, and J. M. Stuart. The cancer genome atlas pan-cancer analysis project. *Nat Genet*, 45(10):1113–20, 2013. (Cited on page 56.)

S. Oltean and D. O. Bates. Hallmarks of alternative splicing in cancer. *Oncogene*, 33(46):5311–5318, 2014. (Cited on page 62.)

Krushna C Patra and Nissim Hay. The pentose phosphate pathway and cancer. *Trends in biochemical sciences*, 39(8):347–354, 2014. (Cited on page 15.)

Stephanos Pavlides, Diana Whitaker-Menezes, Remedios Castello-Cros, Neal Flomenberg, Agnieszka K Witkiewicz, Philippe G Frank, Mathew C Casimiro, Chenguang Wang, Paolo Fortina, Sankar Addya, et al. The reverse warburg effect: aerobic glycolysis in cancer associated fibroblasts and the tumor stroma. *Cell cycle*, 8(23):3984–4001, 2009. (Cited on page 8.)

Salomé S Pinho and Celso A Reis. Glycosylation in cancer: mechanisms and clinical implications. *Nature Reviews Cancer*, 2015. (Cited on page 7.)

Richard Possemato, Kevin M Marks, Yoav D Shaul, Michael E Pacold, Dohoon Kim, Kıvanç Birsoy, Shalini Sethumadhavan, Hin-Koon Woo, Hyun G Jang, Abhishek K Jha, et al. Functional genomics reveal that the serine synthesis pathway is essential in breast cancer. *Nature*, 476(7360):346–350, 2011. (Cited on page 16.)

A. Purmann, J. Toedling, M. Schueler, P. Carninci, H. Lehrach, Y. Hayashizaki, W. Huber, and S. Sperling. Genomic organization of transcriptomes in mammals: Coregulation and cofunctionality. *Genomics*, 89(5):580–7, 2007. (Cited on pages 2 and 28.)

Lakshmi Raj, Takao Ide, Aditi U Gurkar, Michael Foley, Monica Schenone, Xiaoyu Li, Nicola J Tolliday, Todd R Golub, Steven A Carr, Alykhan F Shamji, et al. Selective killing of cancer cells by a small molecule targeting the stress response to ros. *Nature*, 475 (7355):231–234, 2011. (Cited on page 66.)

Jeffrey C Rathmell, Casey J Fox, David R Plas, Peter S Hammerman, Ryan M Cinalli, and Craig B Thompson. Akt-directed glucose metabolism can prevent bax conformation change and promote growth factor-independent survival. *Molecular and cellular biology*, 23(20):7315–7328, 2003. (Cited on page 22.)

Arina Riabinska, Mathias Daheim, Grit S Herter-Sprie, Johannes Winkler, Christian Fritz, Michael Hallek, Roman K Thomas, Karl-Anton Kreuzer, Lukas P Frenzel, Parisa Monfared, et al. Therapeutic targeting of a robust non-oncogene addiction to prkdc in atm-defective tumors. *Science translational medicine*, 5(189):189ra78–189ra78, 2013. (Cited on page 66.)

S. Ros and A. Schulze. Balancing glycolytic flux: the role of 6-phosphofructo-2-kinase/fructose 2,6-bisphosphatases in cancer metabolism. *Cancer Metab*, 1(1):8, 2013. (Cited on pages 21 and 59.)

Claudio R. Santos and Almut Schulze. Lipid metabolism in cancer. *FEBS Journal*, 279(15):2610–2623, 2012. (Cited on pages 18 and 19.)

Eric T Sawey, Maia Chanrion, Chunlin Cai, Guanming Wu, Jianping Zhang, Lars Zender, Alice Zhao, Ronald W Busuttil, Herman Yee, Lincoln Stein, et al. Identification of a therapeutic strategy targeting amplified fgf19 in liver cancer by oncogenomic screening. *Cancer cell*, 19(3):347–358, 2011. (Cited on page 29.)

John C Schell, Kristofor A Olson, Lei Jiang, Amy J Hawkins, Jonathan G Van Vranken, Jianxin Xie, Robert A Egnatchik, Espen G Earl, Ralph J DeBerardinis, and Jared Rutter. A role for the mitochondrial pyruvate carrier as a repressor of the warburg effect and colon cancer cell growth. *Molecular cell*, 56(3):400–413, 2014. (Cited on page 14.)

A. Schulze and A. L. Harris. How cancer metabolism is tuned for proliferation and vulnerable to disruption. *Nature*, 491(7424):364–373, 2012. (Cited on pages 3, 6, 12, and 22.)

Fabiana Schwartzenberg-Bar-Yoseph, Michal Armoni, and Eddy Karnieli. The tumor suppressor p53 down-regulates glucose transporters glut1 and glut4 gene expression. *Cancer research*, 64(7):2627–2633, 2004. (Cited on page 23.)

GL Semenza. Regulation of metabolism by hypoxia-inducible factor 1. In *Cold Spring Harbor symposia on quantitative biology*, volume 76,

pages 347–353. Cold Spring Harbor Laboratory Press, 2011. (Cited on page 23.)

M. Semon and L. Duret. Evolutionary origin and maintenance of coexpressed gene clusters in mammals. *Mol Biol Evol*, 23(9):1715–23, 2006. (Cited on pages 2, 27, and 28.)

David B Shackelford and Reuben J Shaw. The lkb1–ampk pathway: metabolism and growth control in tumour suppression. *Nature Reviews Cancer*, 9(8):563–575, 2009. (Cited on page 24.)

David B Shackelford, Debbie S Vasquez, Jacqueline Corbeil, Shulin Wu, Mathias Leblanc, Chin-Lee Wu, David R Vera, and Reuben J Shaw. mtor and hif-1α-mediated tumor metabolism in an lkb1 mouse model of peutz-jeghers syndrome. *Proceedings of the National Academy of Sciences*, 106(27):11137–11142, 2009. (Cited on page 24.)

A. K. Sharma and R. König. Metabolic network modeling approaches for investigating the "hungry cancer". *Semin Cancer Biol*, 23(4):227–34, 2013. (Cited on pages 3, 5, 8, 13, and 14.)

Ashwini Kumar Sharma, Roland Eils, and Rainer König. Tumor metabolic reprogramming by copy number co- alterations of proximal enzyme coding and cancer causing genes. *Under review*, 2015. (Cited on pages 1, 31, 45, and 65.)

Hyunsuk Shim, Christine Dolde, Brian C Lewis, Chyi-Sun Wu, Gerard Dang, Richard A Jungmann, Riccardo Dalla-Favera, and Chi V Dang. c-myc transactivation of ldh-a: implications for tumor metabolism and growth. *Proceedings of the National Academy of Sciences*, 94(13):6658–6663, 1997. (Cited on page 23.)

Tomer Shlomi, Tomer Benyamini, Eyal Gottlieb, Roded Sharan, and Eytan Ruppin. Genome-scale metabolic modeling elucidates the role of proliferative adaptation in causing the warburg effect. *PLoS Comput Biol*, 7(3):e1002018, 2011. (Cited on page 17.)

G. A. Singer, A. T. Lloyd, L. B. Huminiecki, and K. H. Wolfe. Clusters of co-expressed genes in mammalian genomes are conserved by natural selection. *Mol Biol Evol*, 22(3):767–75, 2005. (Cited on pages 2 and 28.)

N. L. Solimini, Q. Xu, C. H. Mermel, A. C. Liang, M. R. Schlabach, J. Luo, A. E. Burrows, A. N. Anselmo, A. L. Bredemeyer, M. Z. Li, R. Beroukhim, M. Meyerson, and S. J. Elledge. Recurrent hemizygous deletions in cancers may optimize proliferative potential. *Science*, 337(6090):104–9, 2012. (Cited on pages 2, 26, 28, 29, and 65.)

Jaekyoung Son, Costas A Lyssiotis, Haoqiang Ying, Xiaoxu Wang, Sujun Hua, Matteo Ligorio, Rushika M Perera, Cristina R Ferrone,

Edouard Mullarky, Ng Shyh-Chang, et al. Glutamine supports pancreatic cancer growth through a kras-regulated metabolic pathway. *Nature*, 496(7443):101–105, 2013. (Cited on pages 18 and 23.)

Duncan Sproul, Nick Gilbert, and Wendy A Bickmore. The role of chromatin structure in regulating the expression of clustered genes. *Nature Reviews Genetics*, 6(10):775–781, 2005. (Cited on page 26.)

V Stambolic, D MacPherson, D Sas, Y Lin, B Snow, Y Jang, S Benchimol, and TW Mak. Regulation of pten transcription by p53. *Molecular cell*, 8(2):317–325, 2001. (Cited on page 23.)

An S Tan, James W Baty, Lan-Feng Dong, Ayenachew Bezawork-Geleta, Berwini Endaya, Jacob Goodwin, Martina Bajzikova, Jaromira Kovarova, Martin Peterka, Bing Yan, et al. Mitochondrial genome acquisition restores respiratory function and tumorigenic potential of cancer cells without mitochondrial dna. *Cell metabolism*, 21(1):81–94, 2015. (Cited on page 17.)

Annelyse Thévenin, Liat Ein-Dor, Michal Ozery-Flato, and Ron Shamir. Functional gene groups are concentrated within chromosomes, among chromosomes and in the nuclear space of the human genome. *Nucleic acids research*, page gku667, 2014. (Cited on page 28.)

Ines Thiele, Neil Swainston, Ronan MT Fleming, Andreas Hoppe, Swagatika Sahoo, Maike K Aurich, Hulda Haraldsdottir, Monica L Mo, Ottar Rolfsson, Miranda D Stobbe, et al. A community-driven global reconstruction of human metabolism. *Nature biotechnology*, 31(5):419–425, 2013. (Cited on page 8.)

M. A. Thomas, B. Weston, M. Joseph, W. Wu, A. Nekrutenko, and P. J. Tonellato. Evolutionary dynamics of oncogenes and tumor suppressor genes: higher intensities of purifying selection than other genes. *Mol Biol Evol*, 20(6):964–8, 2003. (Cited on page 62.)

Nathan D Trinklein, Shelley Force Aldred, Sara J Hartman, Diane I Schroeder, Robert P Otillar, and Richard M Myers. An abundance of bidirectional promoters in the human genome. *Genome research*, 14(1):62–66, 2004. (Cited on page 27.)

Sevin Turcan, Daniel Rohle, Anuj Goenka, Logan A Walsh, Fang Fang, Emrullah Yilmaz, Carl Campos, Armida WM Fabius, Chao Lu, Patrick S Ward, et al. Idh1 mutation is sufficient to establish the glioma hypermethylator phenotype. *Nature*, 483(7390):479–483, 2012. (Cited on page 25.)

Olesya A Ulanovskaya, Andrea M Zuhl, and Benjamin F Cravatt. Nnmt promotes epigenetic remodeling in cancer by creating a metabolic methylation sink. *Nature chemical biology*, 9(5):300–306, 2013. (Cited on page 22.)

CWM Van Veelen, H Verbiest, GEJ Staal, and AMC Vlug. Alanine inhibition of pyruvate kinase in gliomas and meningiomas: A diagnostic tool in surgery for gliomas? *The Lancet*, 310(8034):384–385, 1977. (Cited on page 25.)

M. G. Vander Heiden, J. W. Locasale, K. D. Swanson, H. Sharfi, G. J. Heffron, D. Amador-Noguez, H. R. Christofk, G. Wagner, J. D. Rabinowitz, J. M. Asara, and L. C. Cantley. Evidence for an alternative glycolytic pathway in rapidly proliferating cells. *Science*, 329(5998): 1492–9, 2010. (Cited on pages 14, 24, and 62.)

Matthew G Vander Heiden, Lewis C Cantley, and Craig B Thompson. Understanding the warburg effect: the metabolic requirements of cell proliferation. *science*, 324(5930):1029–1033, 2009. (Cited on page 17.)

J. H. Vogel, A. von Heydebreck, A. Purmann, and S. Sperling. Chromosomal clustering of a human transcriptome reveals regulatory background. *BMC Bioinformatics*, 6:230, 2005. (Cited on page 27.)

B. Vogelstein, N. Papadopoulos, V. E. Velculescu, S. Zhou, Jr. Diaz, L. A., and K. W. Kinzler. Cancer genome landscapes. *Science*, 339 (6127):1546–58, 2013. (Cited on pages 5, 28, and 35.)

Karen H Vousden and Kevin M Ryan. p53 and metabolism. *Nature Reviews Cancer*, 9(10):691–700, 2009. (Cited on page 24.)

Jing Wang, Jun Qian, Megan D Hoeksema, Yong Zou, Allan V Espinosa, SM Jamshedur Rahman, Bing Zhang, and Pierre P Massion. Integrative genomics analysis identifies candidate drivers at 3q26-29 amplicon in squamous cell carcinoma of the lung. *Clinical cancer research*, 19(20):5580–5590, 2013. (Cited on page 29.)

Otto Warburg et al. On the origin of cancer cells. *Science*, 123(3191): 309–314, 1956. (Cited on page 17.)

Patrick S Ward and Craig B Thompson. Signaling in control of cell growth and metabolism. *Cold Spring Harbor perspectives in biology*, 4 (7):a006783, 2012. (Cited on pages 12, 21, and 22.)

Joe H Ward Jr. Hierarchical grouping to optimize an objective function. *Journal of the American statistical association*, 58(301):236–244, 1963. (Cited on page 43.)

Justin Washtell and Katja Markert. A comparison of windowless and window-based computational association measures as predictors of syntagmatic human associations. In *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing: Volume 2 - Volume 2*, EMNLP '09, pages 628–637, Stroudsburg, PA, USA, 2009. Association for Computational Linguistics. ISBN 978-1-932432-62-6. (Cited on page 39.)

Wu Wei, Vicent Pelechano, Aino I. JÃ€rvelin, and Lars M. Steinmetz. Functional consequences of bidirectional promoters. *Trends in Genetics*, 27(7):267 – 276, 2011. (Cited on page 27.)

F. Wessely, M. Bartl, R. Guthke, P. Li, S. Schuster, and C. Kaleta. Optimal regulatory strategies for metabolic pathways in escherichia coli depending on protein costs. *Mol Syst Biol*, 7:515, 2011. (Cited on pages 36 and 67.)

David R Wise, Ralph J DeBerardinis, Anthony Mancuso, Nabil Sayed, Xiao-Yong Zhang, Harla K Pfeiffer, Ilana Nissim, Evgueni Daikhin, Marc Yudkoff, Steven B McMahon, et al. Myc regulates a transcriptional program that stimulates mitochondrial glutaminolysis and leads to glutamine addiction. *Proceedings of the National Academy of Sciences*, 105(48):18782–18787, 2008. (Cited on pages 18 and 23.)

Amparo Wolf, Sameer Agnihotri, Johann Micallef, Joydeep Mukherjee, Nesrin Sabha, Rob Cairns, Cynthia Hawkins, and Abhijit Guha. Hexokinase 2 is a key mediator of aerobic glycolysis and promotes tumor growth in human glioblastoma multiforme. *The Journal of experimental medicine*, 208(2):313–326, 2011. (Cited on page 14.)

Wen Xue, Thomas Kitzing, Stephanie Roessler, Johannes Zuber, Alexander Krasnitz, Nikolaus Schultz, Kate Revill, Susann Weissmueller, Amy R Rappaport, Janelle Simon, et al. A cluster of cooperating tumor-suppressor gene candidates in chromosomal deletions. *Proceedings of the National Academy of Sciences*, 109(21):8212–8217, 2012. (Cited on page 29.)

T. Yamada, I. Letunic, S. Okuda, M. Kanehisa, and P. Bork. ipath2.0: interactive pathway explorer. *Nucleic Acids Res*, 39(Web Server issue):W412–5, 2011. (Cited on page 36.)

Hui Yang, Dan Ye, Kun-Liang Guan, and Yue Xiong. Idh1 and idh2 mutations in tumorigenesis: mechanistic insights and clinical perspectives. *Clinical Cancer Research*, 18(20):5562–5571, 2012. (Cited on page 25.)

Mary Q Yang, Laura M Koehly, and Laura L Elnitski. Comprehensive annotation of bidirectional promoters identifies co-regulation among breast and ovarian cancer genes. *PLoS Comput Biol*, 3(4):e72, 04 2007. (Cited on page 29.)

Haoqiang Ying, Alec C Kimmelman, Costas A Lyssiotis, Sujun Hua, Gerald C Chu, Eliot Fletcher-Sananikone, Jason W Locasale, Jaekyoung Son, Hailei Zhang, Jonathan L Coloff, et al. Oncogenic kras maintains pancreatic tumors through regulation of anabolic glucose metabolism. *Cell*, 149(3):656–670, 2012. (Cited on page 23.)

Jihye Yun, Carlo Rago, Ian Cheong, Ray Pagliarini, Philipp Angenendt, Harith Rajagopalan, Kerstin Schmidt, James KV Willson, Sandy Markowitz, Shibin Zhou, et al. Glucose deprivation contributes to the development of kras pathway mutations in tumor cells. *Science*, 325(5947):1555–1559, 2009. (Cited on page 23.)

T. I. Zack, S. E. Schumacher, S. L. Carter, A. D. Cherniack, G. Saksena, B. Tabak, M. S. Lawrence, C. Z. Zhang, J. Wala, C. H. Mermel, C. Sougnez, S. B. Gabriel, B. Hernandez, H. Shen, P. W. Laird, G. Getz, M. Meyerson, and R. Beroukhim. Pan-cancer patterns of somatic copy number alteration. *Nat Genet*, 45(10):1134–1140, 2013. (Cited on pages 32, 33, 45, 53, and 66.)

Nousheen Zaidi, Ines Royaux, Johannes V Swinnen, and Karine Smans. Atp citrate lyase knockdown induces growth arrest and apoptosis through different cell-and environment-dependent mechanisms. *Molecular cancer therapeutics*, 11(9):1925–1935, 2012. (Cited on page 20.)

Kathrin Zaugg, Yi Yao, Patrick T Reilly, Karuppiah Kannan, Reza Kiarash, Jacqueline Mason, Ping Huang, Suzanne K Sawyer, Benjamin Fuerth, Brandon Faubert, et al. Carnitine palmitoyltransferase 1c promotes cell survival and tumor growth under conditions of metabolic stress. *Genes & development*, 25(10):1041–1051, 2011. (Cited on page 24.)

Lars Zender, Mona S Spector, Wen Xue, Peer Flemming, Carlos Cordon-Cardo, John Silke, Sheung-Tat Fan, John M Luk, Michael Wigler, Gregory J Hannon, et al. Identification and validation of oncogenes in liver cancer using an integrative oncogenomic approach. *Cell*, 125(7):1253–1267, 2006. (Cited on page 29.)

M. Zhao, J. Sun, and Z. Zhao. Tsgene: a web resource for tumor suppressor genes. *Nucleic Acids Res*, 41(Database issue):D970–6, 2013. (Cited on page 35.)

COLOPHON

This document was typeset using the typographical look-and-feel `classicthesis` developed by André Miede [1]. The style was inspired by Robert Bringhurst's seminal book on typography *"The Elements of Typographic Style"*. LaTeX and LyX was used for the prparation of this document. `classicthesis` was first ported to LyX by Nick Mariette and Ivo Pletikosić [2] created the LyX layout file.

Final Version as of November 9, 2015 (`classicthesis` version 1.0).

1 http://code.google.com/p/classicthesis/
2 http://soundsorange.net/resources/classic-thesis-for-lyx/