Copy Number Alterations in Enzyme-Coding and Cancer-Causing Genes Reprogram Tumor Metabolism

Ashwini Kumar Sharma1,2, Roland Eils2,3, and Rainer König1,2,4

Abstract

Somatic copy number alterations frequently occur in the cancer genome affecting not only oncogenic or tumor suppressive genes, but also passenger and potential codriver genes. An intrinsic feature resulting from such genomic perturbations is the deregulation in the metabolism of tumor cells. In this study, we have shown that metabolic and cancer-causing genes are unexpectedly often proximally positioned in the chromosome and share loci with coamplified copy numbers across multiple cancers (19 cancer types from The Cancer Genome Atlas). We have developed an analysis pipeline, Identification of Metabolic Cancer Genes (iMetCG), to infer the functional impact on metabolic remodeling from such coamplifications and codeletions and delineate genes driving cancer metabolism from those that are neutral. Using our identified metabolic genes, we were able to classify tumors based on their tissue and developmental origins. These metabolic genes were similar to known cancer genes in terms of their network connectivity, isoform frequency, and evolutionary features. We further validated these identified metabolic genes by (i) using gene essentiality data from several tumor cell lines, (ii) showing that these identified metabolic genes are strong indicators for patient survival, and (iii) observing a significant overlap between our identified metabolic genes and known cancer-metabolic genes. Our analyses revealed a hitherto unknown generic mechanism for large-scale metabolic reprogramming in cancer cells based on linear gene proximities between cancer-causing and -metabolic genes. We have identified 119 new metabolic cancer genes likely to be involved in rewiring cancer cell metabolism.

Introduction

Tumorigenesis can occur through numerous mechanisms giving the malignant cells ceaseless proliferative advantages (1). Evolutionary processes result in a perpetual conflict between the fitness of the host cell community and the cancer-initiating cells. Evolutionary forces may have shaped inherent vulnerabilities in the host genome, thus providing 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 putative or tumorigenic or tumor-suppressive mechanisms. The existence of tumor suppressors and oncogenes in the genome makes this evolutionary dilemma apparent (2).

Several studies have shown that the arrangement of genes in the chromosomes (gene order) is not random in eukaryotes and is subjected to natural selection driven by functionality (3–5). Coexpressed genes, in particular, housekeeping genes cluster together in the chromosomes and are often from the same biochemical pathways (6–9). Accordingly, these gene clusters can be coregulated (10, 11) and genomic regions harboring higher density of cancer genes could be under cotranscriptional control (12).

In cancer cells, somatic copy number alteration (SCNA) can affect large segments of the genome and these susceptible regions often contain 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-driving gene. Interestingly, these coaffected genes are now considered to contribute toward cancer potency via cumulative changes to allelic copy numbers (13, 14). We would miss information about these secondary effectors in analyses focusing solely on high-frequency SCNA events, however, taking into account collective gene dosage effects from multiple genes may reveal that a complete cellular process is deregulated. Recent studies have also reported metabolism-specific SCNA-driven functional coalterations between proximal cancer-causing and metabolic genes in renal cell carcinomas, glioblastomas, and breast cancers (15, 16). For example, codeletion of MTA1 with the CDKN2A/2B tumor suppressor gene is a well-known case that occurs across multiple cancer types affecting the biochemical process of adenosine and methionine salvage (17, 18).

Applying the concepts discussed above, we have systematically explored across a wide range of tumor types, the consequences of intrachromosomal linear proximity (and concurrent copy
number coalterations) between metabolic and cancer-causing genes in tumor metabolism. On the basis of our prior knowledge that altered metabolism is crucial to carcinogenesis (19–21), we assessed the functional relevance of such events and elucidated their potential oncogenic effects on tumor metabolic remodeling.

Materials and Methods
Assembly of different datasets
This study entailed integrating mainly two kinds of data sources (i) cancer genomics data (obtained from The Cancer Genome Atlas, TCGA) consisting of transcriptomics and SCNA datasets along with their clinical information from multiple cancer types (Supplementary Fig. S1A; Supplementary Table S1) and (ii) a priori-defined biologic knowledge. This consisted of annotated chromosomal gene positions (from Ensembl Biomart, GrCh38 release), lists of known metabolic genes (from Kyoto Encyclopedia of Genes and Genomes), cancer-causing genes [from four different sources: (i) Cancer gene census (http://cancer.sanger.ac.uk/census/), (ii) tumor suppressor and Oncogenes Explorer (14), (iii) TSGene database (22), and (iv) the list of cancer genes published in ref. 23; see Supplementary Fig. S1B], experimentally identified essential genes [from four different sources (i) Database for Essential Genes (24), (ii) Online Gene Essentiality Database (25), (iii) the list of essential genes from ref. 26 and (iv) ref. 27; see Supplementary Fig. S1C], predicted bottleneck enzymatic genes (provided by the authors of ref. 28), and various gene sets [gene ontology (GO) terms and a list of transcription factors from DBD-transcription factor prediction database; ref. 29]. We integrated these data sources in an analytic framework (described in the next section) to study the dependency of metabolic remodeling in tumors on SCNA and linear gene proximities. Additional details regarding these data sources are given in the Supplementary Methods.

The Identification of Metabolic Cancer Genes analysis pipeline
We set up the Identification of Metabolic Cancer Genes (iMetCG) analysis pipeline to identify proximal cancer–metabolic gene pairs having significantly correlated copy number coalteration and gene coexpression patterns. These gene pairs were further filtered using a priori-defined metabolic functionality features to identify metabolic genes having putative cancer-relevant roles (Fig. 1). The pipeline performs a two-step process consisting of statistical and functional prioritizations. In the first step, the information content for copy number coalterations was calculated for every possible gene pair in the 22 autosomes form each cancer type ($N = 19$). The information content was measured using mutual information such that higher mutual information corresponds to a higher frequency of copy number coalteration. Using the distribution of calculated information content values, we stringently filtered for distinctively coalterated gene pairs from each cancer type satisfying (i) $>99\%$ quantile of the score distribution and (ii) an absolute cooccurrence probability of $>5\%$ (to avoid the low frequency problem, see Supplementary Methods) and (iii) gene pair distance of $<5$ Mb. Following the statistical prioritization step, we selected from each cancer type, only the subset of cancer–metabolic gene pairs from the list of all significantly coalterated gene pairs for our downstream analysis. Next, we identified cancer-metabolic gene pairs for which the expression of the metabolic gene differed significantly ($P \leq 0.05$, Benjamini–Hochberg corrected) between the two groups of tumor samples harboring and lacking the copy number alteration (done for both the metabolic and the cancer gene separately) specific for the cancer type and cancer–metabolic gene pair. These gene pairs will henceforth be referred to as "prioritized cancer-metabolic gene pairs" and the unique set of individual metabolic genes irrespective of their coalterated cancer gene partner will be called the "core metabolic gene set".

The pipeline further filtered the prioritized cancer-metabolic gene pairs by integrating features relevant to cellular functionality like gene essentiality and the regulation of metabolic fluxes using a priori-defined gene lists of essential and bottleneck metabolic genes. In addition, we created a gene-centric network of metabolic genes by applying an analogous methodology used in a previous study (30). Using this network, we identified those genes whose nearest neighbors (NN) in the network were enriched for differential expression ($P \leq 0.05$ and $N \geq 5$) between tumors with diploid (nonaltered) and altered copy numbers of the proximal cancer genes. We term these "predicted metabolic cancer genes," identified after multiple rounds of prioritizations. For technical details regarding the iMetCG pipeline, mathematical description for mutual information and construction of the gene-centric metabolic network, see Supplementary Methods.

Statistical analysis
All $P$ values were calculated using one-sided Wilcoxon test unless otherwise specified. Two-sided Welch $t$ tests were used for differential gene expression analysis in functional prioritization.
steps (coexpression and network analysis). Enrichment tests were performed using one-sided Fisher exact tests. Hierarchical clustering was performed using the Ward linkage method. We used the `dist` and `hclust` base functions available in R. All statistical data analysis was carried out using the R statistical programming environment (www.r-project.org, version 3.0.0).

**Results**

**Cancer–metabolic gene pairs are closer together in the genome than cancer–nonmetabolic gene pairs**

Linear intrachromosomal distances between every protein-coding gene pair midpoints were calculated in each of the 22 autosomes (for the frequency distribution of cancer-causing genes, metabolic genes and the corresponding gene pairs in each chromosome, see Supplementary Fig. S2A and S2B). We used a 1 Mb cutoff to define proximal gene pairs because it was recently shown in a pan-cancer SCNA study that the average size of focal SCNAS away from the telomeres in both amplification and deletion events is less than 1 Mb (31). We observed that the distances between cancer–metabolic gene pairs were significantly smaller \( (P = 0.023, \text{median distance difference} = 33 \text{ Kb}) \) than those between cancer–nonmetabolic gene pairs (Fig. 2A). Furthermore, we verified that the shorter separation between cancer and metabolic genes is unique to the class of metabolic genes by performing this comparison with other functional gene sets having direct relevance to oncogenesis (Fig. 2A). In addition, to assess whether the shorter separation between cancer–metabolic gene pairs did not occur by random chance, we shuffled the gene labels 10,000 times keeping the cancer gene position, chromosomal gene density, and chromosomal affiliations constant (details on the permutation analysis in Supplementary Methods). We recomputed the median distance difference between cancer–metabolic and cancer–nonmetabolic gene pairs using a 1 Mb cutoff in these randomized genomes, and found significant differences between them in the observed and random genomes \( (P = 0.01, \text{Fig. 2B}) \). Interestingly, 63% of all metabolic genes reside within 1 Mb from cancer genes and 83% of all metabolic pathways have at least 50% of their genes within the same proximity. We also noticed a high coverage (despite larger pathway sizes) for several cancer-relevant metabolic pathways, such as oxidative phosphorylation (OXPHOS), purine and pyrimidine metabolism, glycolysis and synthesis of glycerophospholipids, and amino acids (Supplementary Fig. S2C). From these observations, we conclude that metabolic genes are distinctively closer to cancer genes than nonmetabolic genes.

**Cancer–metabolic gene pairs have higher copy number coalterations than cancer–nonmetabolic gene pairs**

We investigated whether the proximity between cancer and metabolic genes also results in higher occurrences of cancer–metabolic gene copy number coalteration. We compared the cooccurrence 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

![Figure 2.](attachment:figure2.png)

Comparison of distance proximities and cooccurrences of copy number alterations between cancer–metabolic and cancer–nonmetabolic gene pairs. A, comparison of intrachromosomal distances between all cancer–metabolic (CG-MG) and cancer–nonmetabolic (CG-NMG) gene pairs separated by <1 Mb distance (above the black bar) followed by the same distance comparison between cancer-causing genes and genes from the GO terms immune processes (IM, GO:0002376), biologic regulation (BR, GO:0065007), signaling process (SG, GO:0005202), and transcription factors (TF, from the DBD-transcription factor prediction database). B, permutation analysis using 10,000 random genomes (10K). The histogram shows the distribution of median distance differences between CG-MG and CG-NMG pairs in the random genomes; the black dashed line represents median distance difference (33 Kb) in the real genome. C, comparison of copy number coalterations (measured as information content) between all CG-MG and CG-NMG pairs from 19 cancer types. D, copy number coalterations between CG-MG and CG-NMG pairs separated by similar distances up to 1 Mb using a sliding window of 250 Kb. The difference in the first bin is not significant.
selected a 1 Mb cutoff and a minimum of 5% absolute cooccurrence frequency to observe relevant SCNA effects. Coalteration frequency for all selected cancer–metabolic and cancer–nonmetabolic gene pairs was measured using normalized pointwise mutual information (npMI). A high npMI value corresponds to a high information content of cooccurring alterations, thus a highly nonrandom event (see Supplementary Methods). We observed significantly higher coalteration frequencies for cancer–metabolic gene pairs \((P = 10^{-9}, \text{Fig. 2C})\), which is in accordance to the smaller cancer–metabolic gene distances that we observed in the previous section. It is however expected that proximal genes in a SCNA-susceptible locus have higher chances of coalteration, thus, we controlled for this confounder by grouping cancer–metabolic and cancer–nonmetabolic gene pairs into categories of similar distance separations (distance bins of \(0–250\) Kb, \(250–500\) Kb, \(500–750\) Kb, and \(750–1,000\) Kb) and calculated the coalteration frequencies individually in each of these bins. Interestingly, again we observed significant differences \((P \leq 0.05, \text{Fig. 2D})\), in all bins except the first (smallest bin). These observations suggest that higher rates of coalterations between cancer–metabolic genes pairs are not simply trivial observations originating because of the distance proximity between them. Therefore, we propose that these coalterations might not just be collateral effects of targeting the cancer-causing gene, but could be functional for at least a subset where the affected metabolic gene produces a phenotype that is beneficial (oncogenes) or detrimental (tumor suppressors) to tumors.

Prioritizing copy number coalterated cancer–metabolic gene pairs using the iMetCG analysis pipeline

We applied the iMetCG analysis pipeline to prioritize cancer–metabolic gene pairs, which were involved in crucial metabolic reprogramming processes leading to functional consequences in tumorigenesis (Fig. 1, see Materials and Methods). First, we identified all significantly coalterated (by copy number) gene pairs in each cancer type using mutual information (gene pair counts in Supplementary Fig. S2D and information content score distribution in Supplementary Fig. S3). These measured information content values were positively correlated with absolute cooccurrence frequencies (Supplementary Fig. S4). 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. Next, we identified the prioritized cancer–metabolic gene pairs by correlating the corresponding copy number and expression data for each pair (associated statistics in Supplementary Table S2). In total, 243 core metabolic genes were identified using the iMetCG pipeline, of which 201 were associated with copy number amplifications and 43 were associated with deletions. Interestingly, coamplification events were more common than codeletion events in these identified pairs, suggesting that tumors prefer inducing metabolic gene expression through copy number amplifications rather than reducing it by deletion events. The list of prioritized cancer–metabolic gene pairs was further reduced by integrating various metabolic functionality features, which resulted in a unique set of 119 predicted metabolic cancer genes (Supplementary Table S3). These metabolic genes comprised of 34 genes identified from the network analysis, 78 essential genes, and 31 bottleneck metabolic genes (Supplementary Fig. S5A). The iMetCG analysis is essentially a stepwise prioritization process that started off with 89,889 unique cancer–metabolic gene pairs (>100,000 individual pairs) and identified 119 putative metabolic cancer genes (Supplementary Fig. S5B). These identified genes consist of novel and known metabolic cancer genes very likely to be involved in metabolic reprogramming of tumor cells.

Identification of synergistic cancer–metabolic gene clusters

To elucidate similarities in somatic cancer progression based on metabolic aberrations, we performed unsupervised hierarchical clustering on the prioritized cancer–metabolic gene pairs identified in at least 2 of 16 cancer types (Fig. 3A, left heatmap, complete heatmap with gene pair information in Supplementary Fig. S6). These prioritized and coexpressed cancer–metabolic gene pairs include many well-known tumor suppressors and oncogenes, which were identified in recent pan-cancer SCNA studies (31, 32), suggesting that common SCNA events occurring in the cancer genomes often target proximal metabolic genes (Table 1). In a second clustering analysis, we selected the core metabolic genes affected in at least 2 cancers \((N = 113; \text{Fig. 3A, right heatmap})\), shows metabolic genes altered in at least 25% of cancer types, complete heatmap in Supplementary Fig. S7). These clustering analyses show that many metabolic genes are affected by similar metabolic rewiring programs across tumor types. In addition, we also observed copy number coalterated genomic hotspots harboring large sets of proximal metabolic and cancer genes (within <1 Mb) across different cancer types. Among others, we identified two such SCNA-susceptible genomic clusters of cancer and metabolic genes that occurred at high coalteration frequencies across several tumors, (i) on chromosome 1, between \(q21.2\) and \(q23.3\) for amplification and (ii) on chromosome 8, between \(p22\) and \(p21.3\) for deletion (Supplementary Fig. S8A). To note, such genomic regions, if targeted by a single SCNA presumably affecting a cancer-driving gene, would collaterally affect a large number of proximal metabolic genes, which may have implications on functionality. Analysis of the chromosomal frequency distribution of these coaltered pairs revealed that most of these coalteration events arise from just few autosomes (Supplementary Fig. S8B). This suggests an intrinsic genome-dependent mechanism, which results in higher rates of SCNA coalteration susceptibility for some chromosomes irrespective of the cancer type.

Tumor classification based on tissue and developmental origins

Cancers are traditionally classified in clinical settings by their histology and/or developmental origins (33). Recent pan-cancer efforts, however, have tried to stratify cancers into their tissues of origin using generic information like SCNA (32, 34). We were able to group tumors into clusters that fit their previously described developmental/tissue origins using the core metabolic gene set altered in at least 2 cancer types \((N = 113)\) identified using the iMetCG pipeline (Fig. 3B). In addition, we also identified novel cluster associations. We reidentified the group of lung adenocarcinomas, urothelial bladder cancers, and the group of head and neck as well as lung squamous cell carcinomas, as reported elsewhere (34). Our analysis correctly clustered cervical cell cancers, a squamous cell carcinoma not included in the previous study, with other squamous cell cancers. Our analyses further classified tumors into mesenchymal-like clusters containing ovarian cancers, endometrial carcinomas, and stomach adenocarcinomas, the first two of which have mesodermal origins. Stomach adenocarcinomas, though not of mesodermal lineage,
are known to undergo epithelial–mesenchymal transition during tumorigenesis (35), which may justify their clustering with this group. Cancers from ecto- and endodermal lineages clustered into three groups: (i) from organ epithelium, including hepatocellular carcinomas and breast cancer, (ii) from surface epithelium, including colorectal and prostate adenocarcinomas, and (iii) of neuroectodermal origin including glioblastomas, lower grade gliomas, and cutaneous melanomas. Surprisingly, upon performing similar clustering analysis using only the coaltered cancer genes yielded considerably weaker clustering results (Supplementary Fig. S9), suggesting that the core metabolic gene signature contains cancer type–specific information. 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 their histologic and developmental origins. This is remarkable considering that a small signature of 113 metabolic genes is sufficient for this physiologically complex class separation.

Copy number coalterations of cancer-causing and -metabolic genes affect crucial metabolic pathways

To identify biochemical pathways affected across cancer types, we mapped the core metabolic gene onto a global human metabolic network. Many of the identified metabolic genes belonged to metabolic processes indispensable for tumor growth like nucleotides, lipids, carbohydrates, amino acids, and energy metabolism pathways (Fig. 4). We focused our interpretations mostly on coamplification events, unless otherwise stated, as these were most commonly observed in our analysis. We identified enzymatic genes from the nucleotide biosynthesis pathway, which converts the monophosphate forms of purines and pyrimidines.

<table>
<thead>
<tr>
<th>Known oncogenes or tumor suppressors</th>
<th>Proximal coaltered metabolic genes</th>
</tr>
</thead>
<tbody>
<tr>
<td>CDKN2A/2B</td>
<td>MTAP, SQLE, NDUFB9</td>
</tr>
<tr>
<td>MYC</td>
<td>DGA71, CYC1, GAPA1, TSTA3, OPLAH, PYCRL</td>
</tr>
<tr>
<td>RECQL4</td>
<td>ABCCC5, PIK3CA, ALG3, MCCCC1, POLR2H, EHHAIDH, B3GNT5, DGKG, NDUFB5</td>
</tr>
<tr>
<td>SOX2</td>
<td>EGFR, CCNE1, FGFR1, SDHC</td>
</tr>
<tr>
<td>EGFR</td>
<td></td>
</tr>
<tr>
<td>CCNE1</td>
<td></td>
</tr>
<tr>
<td>FGFR1</td>
<td></td>
</tr>
<tr>
<td>SDHC</td>
<td></td>
</tr>
</tbody>
</table>

*These tumor suppressor genes and oncogenes are representative examples selected from Fig. 3A.

Figure 3. Identification of commonly coaltered cancer–metabolic gene pairs across cancer types. A, left, binary heatmap (presence, red/green; absence, gray) for all cancer–metabolic (CG-MG) gene pairs identified in at least two cancer types using the iMetCG analysis pipeline after statistical prioritization and coexpression analysis (complete heatmap with gene pair information in Supplementary Fig. S6). Right, heatmap highlighting only the core metabolic gene set (observed in at least 4 of 16 cancers; complete list in Supplementary Fig. S7). Metabolic genes in blue boxes were most commonly copy number coaltered across tumor types. The color bar on the left shows the predicted metabolic cancer genes (black) identified in later steps. B, classification of tumors based on developmental and tissue of origin using the core metabolic gene set (also see Supplementary Fig. S9).
pyrimidines and their deoxy-counterparts to respective di/tri phosphates, as well as some RNA polymerase subunits (Fig. 4, red panel). From our analysis, we noticed that the most frequent codeletion event (in 6/16 cancers) involved the MTAP metabolic enzyme and the CDKN2A/2B tumor suppressor gene (Fig. 3A; Table 1). MTAP is crucial for salvaging adenine and methionine in the cell and this codeletion is a well-known tumorigenic event in multiple cancer types (17, 18). Interestingly, two recent studies have shown that MTAP-depleted tumor cell lines are susceptible to therapeutic intervention by inhibiting PRMT5, an arginine methyltransferase (36, 37). This shows that our methodology can identify such relevant alterations in an unbiased and unsupervised manner.

Several key enzyme-coding genes from glycolysis, tricarboxylic cycle, and fructose/mannose metabolism were also altered across cancer types. Interestingly, the enzyme PFKFB2 is known to produce intermediate metabolites involved in the allosteric control of the glycolytic flux (Fig. 4, blue panel; ref. 38). In addition, we identified genes involved in oxidative phosphorylation belonging to the NADH dehydrogenase complex, cytochrome c oxidase/reductases complex and the ATPase complex. These findings support the accumulating evidence that oxidative phosphorylation is active and functional in many cancers (39). One of the NADH dehydrogenase complex subunit NDUFB9 was among the most frequently copy number coamplified with the MYC oncogene (in 10/16 cancers, Fig. 3A; Table 1).

We also identified several genes involved in glycerophospholipid metabolism (involved in cellular membrane maintenance) whose copy numbers were altered across multiple cancers. Many
of these enzymes were specifically involved in the production of phosphatidic acid, a vital precursor for most glycerolipids (40) and which is also known to regulate mTOR function (Fig. 4, top green panel; ref. 41). We identified several genes from the mevalonate–steroid pathway, which are involved in enzymatic reactions that ultimately produce farnesyl phosphate, an important precursor of anabolic processes (Fig. 4, bottom green panel). Interestingly, the SQLE gene, which encodes the flux-controlling enzyme in the cholesterol–steroid biosynthesis, was one of the two most frequently coalterd metabolic gene with MYC oncogene (in 10/16 cancers, Fig. 3A; Table 1), suggesting an important role of steroid anabolism across tumors types.

In amino acid metabolism, several genes in branched chain amino acid metabolism were affected, which are also involved in fatty acid degradation. These reactions normally support anaerobiosis of the Krebs cycle while other affected genes belonged to serine, proline, and glutaminolysis metabolic pathways (Fig. 4, yellow panel). In total, the iMetCG pipeline identified both known and novel metabolic processes that are vital for cancer metabolism. Besides the identified genes that have been previously shown to be relevant in cancer metabolism (19, 21), the novel genes highlight interesting and novel aspects of metabolic network remodeling that warrant further investigation.

**Predicted metabolic cancer genes share similar features with known cancer genes**

We analyzed our predicted cancer genes on functional and genetic features that are distinctively observed in known cancer-causing genes. It is known that cancer genes are highly connected in cellular networks (42), show a high diversity in their isoforms (43) and are highly conserved during evolution (44). Strikingly, our predicted metabolic cancer genes are very similar in these distinctive features (Fig. 5).

In the human metabolic network, like any other cellular network, disrupting highly connected genes or hubs can have a drastic impact on function (45). Using our gene-centric metabolic network, we compared network connectivity of our predicted metabolic cancer genes with other metabolic genes. We observed that the average connectivity of the former was higher in the network ($P = 0.05$, Fig. 5A), suggesting that their perturbations may cause remodeling of the cancer metabolic landscape.

The second cancer-specific feature analyzed was gene isoform diversity, which is an important determinant of cancer progression and malignancy (43). A well-known example of a differentially expressed metabolic gene isoform in tumor metabolism is the PKM2 isoform, which promotes rapid growth by supporting anabolic metabolism through the regulation of glycolytic flux (46). We investigated the isoform counts of our identified metabolic cancer genes and found that they shared a similar frequency of isoforms with known cancer genes. However, they showed a higher frequency of isoforms compared with the remaining metabolic genes in the genome ($P = 0.05$, Fig. 5B).

Finally, we focused on evolutionary features exhibited by cancer genes, which are known to be highly conserved (44). We investigated whether our identified metabolic cancer genes shared similar selection pressure and evolutionary rates. The ratio of nonsynonymous ($dN$) to synonymous ($dS$) substitutions in orthologous genes of distant species gives a robust measure for selection pressure. $dN/dS < 1$ implies purifying selection (higher conservation), while $dN/dS > 1$ suggests positive selection and lower conservation (47). We used the GRCh38 Ensemble release for human and GRCh38.p2 for mouse from Biomart to obtain $dN/dS$ ratios for human-mouse orthologous genes. It has recently been shown that cancer and metabolic genes are evolutionarily more conserved and under stronger purifying selection than other genes (48). We made similar observations while comparing the distributions of $dN/dS$ ratios for cancer and metabolic genes. We also observed that the predicted metabolic cancer genes share similar median $dN/dS$ ratios with cancer genes, which is lower compared with the remaining metabolic genes ($P = 0.04$, Fig. 5C).

We also compared the evolutionary rates among cancer, metabolic and putative metabolic cancer genes using data calculated elsewhere (48). In concordance to the $dN/dS$ ratios, we observed that our identified metabolic cancer genes showed distinctively lower evolutionary rates than the other metabolic genes ($P = 0.002$, Fig. 5D). On the basis of these cancer relevant features, we find that our predicted metabolic cancer genes were similar to known cancer genes while being distinctly different from the remaining metabolic genes, in line with our hypothesis that these genes are central to cancer cell metabolism.

**Validation of the predicted metabolic cancer genes**

We evaluated the effect of our predicted metabolic cancer genes on tumor cell growth and survival using gene essentiality data measured in cancer cell lines and patient survival data,
respectively. We used a previously published dataset of a genomewide (pooled) shRNA screen for gene essentiality targeting approximately 15,500 genes in 72 different cancer cell lines (breast, \(N = 29\); ovarian, \(N = 15\); pancreatic, \(N = 28\); refs. 49, 50). In a global analysis, we identified all metabolic genes significantly affecting tumor cell survival in \(\geq 2\) cell lines of any tumor type (using \(P \leq 0.05\) cutoff, for details on how significance of cell viability is calculated, see ref. 50). Out of these, we observed that a significant fraction of our predicted metabolic cancer genes (43%) were enriched (\(P = 0.04\), Fig. 6A) in affecting tumor cell survival compared with other metabolic genes. In concordance, we also observed that the average gene knockdown score (GARP scores, for details see ref. 50) was significantly lower (\(P = 0.04\), Fig. 6B) for the predicted metabolic cancer genes compared with the other metabolic genes. In addition, using this dataset, we also performed a cancer-specific analysis for ovarian and breast tumors. Pancreatic tumors were not analyzed as the corresponding patient tumor data was not available in the TCGA data we used in our analysis. We found that 31% and 20% of the predicted metabolic cancer genes specifically identified in ovarian and breast cancers, respectively, reduced tumor growth significantly in at least two cell lines from the same cancer type (for the list of these genes and cell lines, see Supplementary Table S4).

Next, we investigated the role of the predicted metabolic cancer genes on patient survival. We calculated the survival difference (using log-rank test) between cancer samples harboring or lacking copy number alterations. Only those metabolic genes with high-level of copy number alterations (GISTIC2 values of +2 or −2) in at least 5% of the samples were used from each cancer type. Strikingly, the relative percentage of significant survival differences involving our predicted metabolic cancer gene was higher than for other metabolic genes (\(P = 0.002\), Fig. 6C), suggesting that these genes are a crucial determinant of patient survival across the investigated cancer types.

Finally, we evaluated the capability of our analysis pipeline to recover known cancer-causing metabolic genes. We measured the over-representation of the predicted metabolic cancer genes in a list of known cancer-causing metabolic genes and observed that they were significantly higher compared with other metabolic genes (\(P = 0.006\), Fig. 6D). In total, our validation results clearly show that many of our predicted metabolic cancer genes directly affect tumor growth and survival, thus affirming our methodology, which was developed for discovering such cancer-relevant metabolic genes.

**Discussion**

If during the course of genome evolution, genes vital for cellular proliferation were preferentially positioned near cancer "drivers" in the chromosomes or vice versa, then in tumors, these proximally placed genes would be more susceptible to copy number coalteration. Ideally, such a strategy would be beneficial (in case of coamplifications) or detrimental (in case of codeletions) to cancer progression. The nonrandom organization of genes in the chromosomes is a consequence of millions of years of genome evolution. However, the evolutionary trajectory in tumors occurs in much shorter time scales, that is, within the "life span" of the individual carrying the tumor, but is still governed by the same evolutionary principles. We propose that those tumor cells that can take advantage of the genomic colocalization between cancer and metabolic genes increase their fitness and spread in the tumor cell population.

In this study, we find that metabolic genes and cancer-causing genes are distinctively close to each other in the human genome. To understand the implications of this genomic proximity, we developed the iMetCG analysis pipeline and applied it to 19 different cancer types to identify novel metabolic genes involved in carcinogenesis. In addition, to overcome the difficulty in interpreting copy number coalteration among spatially proximal genes, we performed further steps for functional prioritization. This involved correlating expression and SCNA data and exploiting a priori knowledge of metabolic functionality features to distinguish the relevant copy number coaltered cancer–metabolic gene pairs from artifacts.

Indeed, the metabolic genes identified using the iMetCG pipeline target many metabolic pathways indispensable to cancer proliferation. Furthermore, these genes could successfully group cancers according to their histology. This observation has significant clinical relevance, as a metabolic gene-specific drug could be effectively used in multiple cancers of similar origin and would presumably have comparable drug-induced effects. Also, our predicted metabolic cancer genes shared similar characteristics with known cancer genes, affected tumor growth, and patient survival thus directly affirming the functional roles of these genes.
in neoplasms. However, further experimental validations and mechanistic studies will be required to identify actionable targets.

We have generated a novel resource for cancer-associated metabolic genes, derived from proximal cancer–metabolic gene pairs affected by copy number alterations in a large set of cancer types. These identified metabolic cancer genes serve the atypical metabolic needs of tumor cells and are vital to cancer progression. We believe this computationally derived resource will be useful to the cancer community to explore the roles of unconventional metabolic genes in tumor formation.

Disclosure of Potential Conflicts of Interest
No potential conflicts of interest were disclosed.

Authors’ Contributions
Conception and design: A.K. Sharma, R. Eils, R. König
Development of methodology: A.K. Sharma, R. König
Acquisition of data (provided animals, acquired and managed patients, provided facilities, etc.): A.K. Sharma
Analysis and interpretation of data (e.g., statistical analysis, biostatistics, computational analysis): A.K. Sharma, R. König
Writing, review, and/or revision of the manuscript: A.K. Sharma, R. Eils, R. König

References
4. Singer GA, Lloyd AT, Huminiecki LB, Wolfe KH. Clusters of co-expressed gene pairs affected by copy number alterations in a large set of cancer types. Metabolic genes, derived from proximal cancer–metabolic gene pairs, are vital to cancer progression. We believe this computationally derived resource will be useful to the cancer community to explore the roles of unconventional metabolic genes in tumor formation.

Acknowledgments
The authors thank the TCGA research network http://cancergenome.nih.gov/ for making the multipletplatform genomic data for various tumor types publicly available on which the results of this study are mainly based. The authors also thank Peter Lichter for his valuable suggestions to their analysis and Kathy Astrahanseff for her stylistic corrections.

Grant Support
A.K. Sharma and R. König were supported by grants from the Federal Ministry of Education and Research (BMBF), Germany (FKZ 0316168D, FKZ 0316076C, FKZ 01EO1502, FKZ 01Z2X1302B and FKZ 01EO1008). A.K. Sharma was also supported by the Helmholtz International Graduate School for Cancer Research Fellowship.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked advertisement in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Received September 8, 2015; revised May 2, 2016; accepted May 11, 2016; published OnlineFirst May 23, 2016.

Advisory editorial advertisement
Published OnlineFirst May 23, 2016; DOI: 10.1158/0008-5472.CAN-15-2350

© 2016 American Association for Cancer Research. cancerres.aacrjournals.org Downloaded from cancerres.aacrjournals.org on April 23, 2021.


Copy Number Alterations in Enzyme-Coding and Cancer-Causing Genes Reprogram Tumor Metabolism

Ashwini Kumar Sharma, Roland Eils and Rainer König


Updated version
Access the most recent version of this article at:
doi:10.1158/0008-5472.CAN-15-2350

Supplementary Material
Access the most recent supplemental material at:
http://cancerres.aacrjournals.org/content/suppl/2016/05/21/0008-5472.CAN-15-2350.DC1

Cited articles
This article cites 50 articles, 12 of which you can access for free at:
http://cancerres.aacrjournals.org/content/76/14/4058.full#ref-list-1

Citing articles
This article has been cited by 3 HighWire-hosted articles. Access the articles at:
http://cancerres.aacrjournals.org/content/76/14/4058.full#related-urls

E-mail alerts
Sign up to receive free email-alerts related to this article or journal.

Reprints and Subscriptions
To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org.

Permissions
To request permission to re-use all or part of this article, use this link http://cancerres.aacrjournals.org/content/76/14/4058.
Click on "Request Permissions" which will take you to the Copyright Clearance Center's (CCC) Rightslink site.