The Warburg effect: persistence of stem-cell metabolism in cancers as a failure of differentiation

M. Riester1,2†, Q. Xu1,2†, A. Moreira3, J. Zheng4, F. Michor1,2,5,6,7* & R. J. Downey8*

1Department of Biostatistics and Computational Biology, Dana-Farber Cancer Institute, Boston; 2Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston; 3Department of Pathology, NYU Medical Center, New York; 4Department of Epidemiology and Biostatistics, Memorial Sloan Kettering Cancer Center, New York; 5Department of Stem Cell and Regenerative Biology, Harvard University, Cambridge; 6Broad Institute of Harvard and MIT, Cambridge; 7Center for Cancer Evolution, Dana-Farber Cancer Institute, Boston; 8Thoracic Service, Department of Surgery, Memorial Hospital, Memorial Sloan Kettering Cancer Center, New York, USA

*Correspondence to: Prof. Franziska Michor, Department of Biostatistics and Computational Biology, Dana-Farber Cancer Institute, Boston, MA 02215, USA. Tel: +1-617-632-50-45; E-mail: michor@jimmy.harvard.edu
Prof. Robert J. Downey, Thoracic Service, Department of Surgery, Memorial Hospital, Memorial Sloan Kettering Cancer Center, New York, NY 10065, USA. Tel: +212-639-8124; E-mail: downeyr@mskcc.org.
†These authors contributed equally to this work.

Background: Two recent observations regarding the Warburg effect are that (i) the metabolism of stem cells is constitutive (aerobic) glycolysis while normal cellular differentiation involves a transition to oxidative phosphorylation and (ii) the degree of glucose uptake of a malignancy as imaged by 18F-fluorodeoxyglucose positron emission tomography (FDG–PET) is associated with histologic measures of tumor differentiation. Combining these observations, we hypothesized that the high levels of glucose uptake observed in poorly differentiated cancers may reflect persistence of the glycolytic metabolism of stem cells in malignant cells that fail to fully differentiate.

Patients and methods: Tumor glucose uptake was measured by FDG–PET in 552 patients with histologically diverse cancers. We used normal mixture modeling to explore FDG–PET standardized uptake value (SUV) distributions and tested for associations between glucose uptake and histological differentiation, risk of lymph node metastasis, and survival. Using RNA-seq data, we carried out pathway and transcription factor analyses to compare tumors with high and low levels of glucose uptake.

Results: We found that well-differentiated tumors had low FDG uptake, while moderately and poorly differentiated tumors had higher uptake. The distribution of SUV for each histology was bimodal, with a low peak around SUV 2–5 and a high peak at SUV 8–14. The cancers in the two modes were clinically distinct in terms of the risk of nodal metastases and death. Carbohydrate metabolism and the pentose-related pathway were elevated in the poorly differentiated/high SUV clusters. Embryonic stem cell-related signatures were activated in poorly differentiated/high SUV clusters.

Conclusions: Our findings support the hypothesis that the biological basis for the Warburg effect is a persistence of stem cell metabolism (i.e. aerobic glycolysis) in cancers as a failure to transition from glycolysis-utilizing undifferentiated cells to oxidative phosphorylation-utilizing differentiated cells. We found that cancers cluster along the differentiation pathway into two groups, utilizing either glycolysis or oxidative phosphorylation. Our results have implications for multiple areas of clinical oncology.

Key words: FDG–PET standardized uptake values, stem cell, metabolism, epithelial malignancies, bimodal distributions

Introduction

In normoxia, fully differentiated tissues primarily utilize oxidative phosphorylation, but when hypoxic, they switch to glycolysis. Conversely, normal embryonic stem cells constitutively utilize glycolysis. In the 1950s, Otto Warburg [1, 2] observed that many cancers also constitutively utilize glycolysis—that is, they are glycolytic in both hypoxic and normoxic environments, hence using ‘aerobic glycolysis’. Because cancers utilizing aerobic glycolysis need increased levels of glucose, they have increased numbers of cell surface glucose transporters. This fact has been exploited in positron emission tomography (PET) based on 18F-fluorodeoxyglucose (FDG), a modified form of glucose so that it can be taken up by...
cells but can neither be metabolized nor excreted, and therefore accumulates intracellularly and can be measured as a standardized uptake value (SUV). The degree of glucose uptake of a malignancy as imaged by FDG–PET is associated with histologic measures of tumor differentiation [3–5]. Based on these observations, we hypothesized that glycolytic cancer cells may have persistence of stem-cell metabolism as a result of a failure of normal differentiation. Here, we investigated the relationship between glucose uptake as measured by PET imaging and patient/tumor characteristics including histologic differentiation and survival as well as the expression patterns of metabolism- and stem-cell-related genes.

Materials and methods

Patient cohorts

A waiver of authorization was obtained from the Memorial Sloan Kettering Cancer Center Institutional Review Board (IRB) to perform this retrospective study. Data including demographic and clinical characteristics were either extracted from prospectively maintained databases (including the Memorial Hospital Thoracic Service database and the Hepatopancreaticobiliary Service databases), or from data collected from previously completed retrospective studies [6, 7]. The datasets aggregated included patients with the following characteristics:

(i) Patients with newly diagnosed and untreated non-small-cell lung cancer (NSCLC) who underwent R$_0$ anatomic resections on whom PET was carried out preoperatively.
(ii) Patients with newly diagnosed and untreated esophageal carcinoma who underwent R$_0$ anatomic resections on whom PET was carried out preoperatively.
(iii) Patients with newly diagnosed and untreated cholangiocarcinoma who underwent R$_0$ resections.
(iv) Patients with NSCLC on whom PET was carried out before and after induction therapy for NSCLC who then underwent R$_0$ resections.
(v) Patients with esophageal cancer on whom PET was carried out before and after induction therapy who then underwent R$_0$ resections.

Tumor pathologic differentiation status was obtained from the pathology report. PET SUV values were derived from radiology reports.

RNA sequencing data

Sample collection was conducted by the Director’s Challenge (DC) project, a consortium of four institutions: University of Michigan Cancer Center (UM), H. Lee Moffitt Cancer Center (HLM), Memorial Sloan-Kettering Cancer Center (MSKCC) and the Dana-Farber Cancer Institute (DFCI). In this study, we utilized data from the 20 MSKCC samples for which SUV data was available. Demographic and clinical information for the DC validation cohort of lung adenocarcinoma patients from MSKCC (the MSKCC cohort) [7, 8] was obtained from the MSKCC Thoracic Service database. One microgram of total RNA (RNA integrity number varying from 6 to 8.9) underwent ribosomal depletion and Truseq library preparation according to instruction provided by Illumina (TruSeq$^\text{Stranded}$ Total RNA LT, cat#RS-122-2202), with six cycles of PCR. Samples were barcoded and run on a HiSeq 2500 in a 50 bp/50 bp paired-end run, using the TruSeq SBS Kit v3 (Illumina). On average, 74 million paired reads were generated per sample, 24% of the data mapped to the transcriptome (range from 5% to 34%). Isoform-level expression of mRNA sequencing data was generated per sample, 24% of the data mapped to the transcriptome (range from 5% to 34%). Isoform-level expression of mRNA sequencing data was generated per sample, 24% of the data mapped to the transcriptome (range from 5% to 34%). Isoform-level expression of mRNA sequencing data was generated per sample, 24% of the data mapped to the transcriptome (range from 5% to 34%). Isoform-level expression of mRNA sequencing data was generated per sample, 24% of the data mapped to the transcriptome (range from 5% to 34%). Isoform-level expression of mRNA sequencing data was generated per sample, 24% of the data mapped to the transcriptome (range from 5% to 34%). Isoform-level expression of mRNA sequencing data was generated per sample, 24% of the data mapped to the transcriptome (range from 5% to 34%). Isoform-level expression of mRNA sequencing data was generated per sample, 24% of the data mapped to the transcriptome (range from 5% to 34%). Isoform-level expression of mRNA sequencing data was generated per sample, 24% of the data mapped to the transcriptome (range from 5% to 34%). Isoform-level expression of mRNA sequencing data was generated per sample, 24% of the data mapped to the transcriptome (range from 5% to 34%). Isoform-level expression of mRNA sequencing data was generated per sample, 24% of the data mapped to the transcriptome (range from 5% to 34%). Isoform-level expression of mRNA sequencing data was generated per sample, 24% of the data mapped to the transcriptome (range from 5% to 34%). Isoform-level expression of mRNA sequencing data was generated per sample, 24% of the data mapped to the transcriptome (range from 5% to 34%).

Normal mixture modeling

PET scores were analyzed with mclust version 4.2, a normal mixture modeling software. A small number of extreme outliers (6 out of 985 PET SUV scores), defined as samples with SUVs larger than four standard deviations above the mean, were excluded for model training only. This approach was taken since mclust tends to place these outliers into separate mixture components (clusters). The optimal number of clusters was defined with the corresponding feature available in mclust for each dataset separately. mclust calculates the Bayesian Information Criterion (BIC) for different models and the model with optimal BIC is chosen; complex models with higher numbers of clusters are only selected when the increase in complexity yields a sufficiently large increase in the goodness of fit of the model. Cluster numbers from one to nine were tested. In all but two cases, the optimal model was bimodal; one dataset (esophageal squamous cell carcinoma) displayed clear bimodal SUV patterns, but the bimodal model was not significantly better than the unimodal model due to the small sample size ($n = 31$). The other dataset (lung adenocarcinoma) had an optimum of three modes, but the first two modes were very similar (PET SUV 1.8 and 3.3). We thus used a bimodal model for all datasets.

We assessed the robustness of model training by bootstrapping: for a dataset with $N$ samples, we randomly drew $N$ samples with replacement and trained a bimodal model. This procedure was repeated 1000 times and the observed variances of the two modes were used to calculate 95% confidence intervals of the two modes. To estimate the robustness of the number of modes, we carried out the same bootstrapping without forcing a bimodal model and reported the percentage of bootstrap replicates for which the best model was bimodal.

We used the wilcox.test R function to test for statistically significant difference in PET SUV and differentiation and node status.

We independently examined the association of PET SUV and node status by using the area under the receiver-operating characteristic curve (AUC) and local regression. The local regression was carried out using the loess R function with default parameters.

We carried out a Kaplan–Meier analysis to determine the survival associated with the participant PET SUV clusters. Corresponding P-values were calculated with the log-rank test, implemented in the survival R package (version 2.37). All statistical tests were two-sided. The raw data of SUV score are provided in Supplementary Table S4 and S5, available at Annals of Oncology online.

Results

We first examined the relationship between histologic differentiation of diverse cancer types and the maximum SUV of the primary site of disease in a unique dataset of 552 patients (Figure 1).
We found that well-differentiated tumors had strikingly low FDG avidity, while moderately and poorly differentiated tumors predominately had high FDG avidity.

The distributions of PET SUVs for a diverse group of untreated epithelial malignancies including cholangiocarcinoma, esophageal adenocarcinoma, esophageal squamous cell carcinoma, lung adenocarcinoma and lung squamous cell carcinoma (Figure 2; supplementary Figure S1, available at Annals of Oncology online) were robustly bimodal. This indicates that there are two kinds of tumors—SUV low and SUV high—and that glucose avidity is not a simple continuous tumor characteristic. Supplementary Table S2, available at Annals of Oncology online, lists the peak

![Figure 1](https://academic.oup.com/annonc/advance-article-abstract/doi/10.1093/annonc/mdx645/4553167)

![Figure 2](https://academic.oup.com/annonc/advance-article-abstract/doi/10.1093/annonc/mdx645/4553167)
SUVs of the SUV modes of these different histologies. The modes were remarkably similar across cancer types treated surgically, with the first mode usually below 6 and the second above 8.

We then examined whether cancers associated with the high or low SUV modes were clinically distinct. Figure 3 displays the risk of lymph node metastases as a function of 18F-FDG PET SUVmax in patients with surgically treated cholangiocarcinoma, esophageal squamous and adenocarcinomas, and lung squamous and adenocarcinomas. In each example, the lower SUV mode was associated with a lower risk and the higher SUV mode a higher risk of lymph node metastases being present (supplementary Figure S2A and B, available at Annals of Oncology online). Also, congruent with previous studies, the high SUV mode tumors were associated with a markedly worse prognosis after resection than the low SUV mode tumors in Kaplan–Meier survival analysis (supplementary Figure S3, available at Annals of Oncology online). We then investigated the distribution of SUVs before and after induction chemotherapy for NSCLC and esophageal carcinoma and found a strong shift between SUV modes after treatment (Figure 4).

Finally, we sought to explore the differences in gene expression of metabolism- and stem-cell-related genes of the two SUVmax modes (well-differentiated/low SUV and moderate-poorly differentiated/high SUV). A heat map combined with regression analysis of metabolism-related pathways for samples from the MSKCC Director’s Challenge dataset is shown in Figure 5. Figure 5A demonstrates a significant separation of the high SUV specimens from the low SUV specimens (P < 0.01, Fisher’s exact test); the left cluster contains 9 of the 13 samples with low SUV (SUV < 6) and the right cluster has 7 of 7 samples with high SUV (SUV > 6). We found that carbohydrate/pentose-related genes were elevated only in the high SUV cluster. The distribution of high and low SUVs appears to be the result of an activation of these pathways in the high SUV cluster. As expected, a proliferation signature (‘Ben-Porath.proliferation’ [11]) correlates with SUV (FDR < 0.001, Figure 5B; supplementary Table S3, available at Annals of Oncology online), implying that the high SUV cluster is characterized by a high proliferation rate. In addition, we observed a statistically significantly activation (P < 0.01) of the Ben-Porath stem-cell signatures [13], including genes overexpressed in human ES cells (Ben-Porath.ES.exp.1 and Ben-Porath.ES.exp.2) and c-Myc targets in ES cells (Ben-Porath.Myc.targets). The PRC2 target signature that is repressed in stem cells [13] was negatively correlated with SUV. Beside these, we also compared signatures of lung adenocarcinoma specimens with that of hMSC. Most signatures overexpressed in hMSC were also activated in high SUV specimens, suggesting that the signature profile of high SUV specimens is close to that of stem cells (supplementary Figure S4, available at Annals of Oncology online).

**Discussion**

When oxygen is readily available, fully differentiated cells generate ATP primarily through mitochondrial oxidative phosphorylation. If the supply of oxygen is limited, differentiated cells utilize glycolysis, a cytoplasmic process not requiring oxygen. If oxygen levels are restored, cells revert to oxidative phosphorylation. In the 1950s, Otto Warburg observed that many cancers utilize ‘aerobic glycolysis’, that is, are glycolytic in both normoxic and hypoxic environment (now commonly referred to as the ‘Warburg effect’). This led to the development of 18F-FDG PET for the imaging of malignant disease [14–17]. Later on it was...
Figure 4. The distribution of SUV\textsubscript{max} pre- and post-induction chemotherapy for esophageal carcinoma (A and B) and non small cell lung carcinoma (C and D).

Figure 5. Heat map combined with regression analysis of metabolism-related pathways for the 20 MSKCC Director’s Challenge samples of lung adenocarcinoma with available PET SUV data. (A) Red represents the low SUV mode, blue the high SUV mode. There is a significant \((P < 0.01, \text{Fisher’s exact})\) separation of the SUV modes where the left cluster contains 9 of the 13 samples with low SUV (SUV < 6.0) and the right cluster has 7 of 7 samples with high SUV (SUV > 6.0). In the heat map, red represents activation of the pathways, and blue represents repression of the pathways. Most of the ES cell-related pathways are activated in the high SUV cluster and display statistically highly significant correlation with SUV\textsubscript{max}. Carbohydrate-/pentose-related pathway is also elevated in the high SUV cluster. (B) Regression analysis for SUV with signature enrichment score of ES cell-related pathways. ***\(P < 0.001\), **\(P < 0.01\), *\(P < 0.05\) (FDR of signature enrichment score and SUV).
found that the degree of FDG avidity of a cancer was prognostic [3, 7, 18–20] but the underlying biology of the Warburg effect has remained obscure.

Here we observed that poorly differentiated tumors have higher SUVs than well-differentiated tumors [3–5] (Figure 1). The differentiation process from stem cells to somatic tissue involves both the gain of markers of differentiation (e.g. production of mucin or cytokeratins) and the loss of stem-cell markers [21]. This led us to hypothesize that the metabolism of cancers may represent persistence of the metabolism of stem cells in cancers, rather than the appropriate switch to oxidative phosphorylation occurring during normal differentiation. Our findings are consistent with the hypothesis that the retained stem-cell metabolism is the basis of the Warburg effect in cancers. First, the degree of histologic differentiation was found to be closely linked to glucose uptake, with poorly differentiated tumors demonstrating high avidity for FDG and the well-differentiated tumors demonstrating low avidity for FDG (Figure 1). Second, using data from PET scans on 552 patients, we demonstrated that SUVs follow a remarkably robust and similar bimodal distribution across cancer types (Figure 2). Third, cancers with a high SUV have a higher likelihood of having nodal metastases (Figure 3). Interestingly, consistent with the bimodal distribution of SUVs, the relationship was discontinuous; tumors with a high glycolytic activity (i.e. above SUV of 8) had a steady risk of finding nodal metastases, but as the SUV fell below 7, there was an abrupt reduction to near zero of the risk of finding nodal metastases (supplementary Figure S2A, available at Annals of Oncology online). Fourth, tumors with a high glycolytic activity (above an SUV of 8) were associated with a worse prognosis than tumors with low SUV (supplementary Figure S3, available at Annals of Oncology online). Finally, a comparison of metabolism- and stem-cell-related gene expression showed that carbohydrate/pentose/nucleotide synthesis-related genes were elevated only in tumors that had high glucose uptake and were similar in gene expression patterns to stem cells (Figure 5).

The modal distribution of cancers has clinical implications. For example, if lung cancers of the high and low SUV modes have different times to recurrence, and different prognoses after the diagnosis of recurrent disease, then it should be possible to model the cost-effectiveness of different schedules of clinical follow-up. Similarly, if lung cancers of the high and low SUV modes differ in levels of aggressiveness, then the time to detection during cancer screening would likely be different. The Early Lung Cancer Action Program [22] which found that 92% of lung adenocarcinomas detected due to prevalence (initial) CT screening exams were poorly differentiated and 8% well differentiated. By comparison, 97% found on incidence (subsequent) screening exams were poorly differentiated and 3% well differentiated.

Finally, we observed that before chemotherapy, the SUVs of lung adenocarcinomas were primarily high and the SUVs after chemotherapy were primarily low (Figure 4), a finding which remains to be explained. It is possible that tumors contain two populations of cells of different metabolisms [23], which also differ in chemosensitivity: a more poorly differentiated subclone utilizing aerobic glycolysis may be more susceptible to chemotherapy, leaving a differentiated oxidative phosphorylation population after treatment. Alternately, a poorly differentiated tumor utilizing aerobic glycolysis may differentiate during chemotherapy into a population of differentiated cells utilizing oxidative phosphorylation [24]. These possibilities remain to be investigated.

Acknowledgements

We acknowledge the use of the MSKCC Integrated Genomics Operation Core and helpful discussions with members of the Michor laboratory.

Funding

NCI Cancer Center Support Grant (CCSG, P30 CA08748), Cycle for Survival (no grant number applies), the Marie-Josée and Henry R. Kravis Center for Molecular Oncology (no grant number applies), and the Dana-Faber Cancer Institute Physical Sciences Oncology Center (U54CA193461).

Disclosure

The authors have declared no conflicts of interest.

References


