



Glutathione peroxidase 2 is a metabolic driver of the tumor immune microenvironment and immune checkpoint inhibitor response

Kazi Mokim Ahmed,¹ Ratna Veeramachaneni,² Defeng Deng,¹ Nagireddy Putluri,³ Vasanta Putluri,⁴ Maria F Cardenas,⁵ David A Wheeler,⁵ William K Decker,^{3,6} Andy I Frederick,⁷ Sawad Kazi,⁸ Andrew G Sikora,² Vlad C Sandulache ,^{1,9,10} Mitchell J Frederick ¹

To cite: Ahmed KM, Veeramachaneni R, Deng D, *et al.* Glutathione peroxidase 2 is a metabolic driver of the tumor immune microenvironment and immune checkpoint inhibitor response. *Journal for ImmunoTherapy of Cancer* 2022;**10**:e004752. doi:10.1136/jitc-2022-004752

► Additional supplemental material is published online only. To view, please visit the journal online (<http://dx.doi.org/10.1136/jitc-2022-004752>).

Accepted 13 June 2022



© Author(s) (or their employer(s)) 2022. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

For numbered affiliations see end of article.

Correspondence to

Dr Mitchell J Frederick; Mitchell.Frederick@bcm.edu

ABSTRACT

Background The existence of immunologically ‘cold tumors’ frequently found across a wide spectrum of tumor types represents a significant challenge for cancer immunotherapy. Cold tumors have poor baseline pan-leukocyte infiltration, including a low prevalence of cytotoxic lymphocytes, and not surprisingly respond unfavorably to immune checkpoint (IC) inhibitors. We hypothesized that cold tumors harbor a mechanism of immune escape upstream and independent of ICs that may be driven by tumor biology rather than differences in mutational neoantigen burden.

Methods Using a bioinformatic approach to analyze TCGA (The Cancer Genome Atlas) RNA sequencing data we identified genes upregulated in cold versus hot tumors across four different smoking-related cancers, including squamous carcinomas from the oral cavity (OSCC) and lung (LUSC), and adenocarcinomas of the bladder (BLCA) and lung (LUAD). Biological significance of the gene most robustly associated with a cold tumor phenotype across all four tumor types, glutathione peroxidase 2 (GPX2), was further evaluated using a combination of in silico analyses and functional genomic experiments performed both in vitro and in vivo with preclinical models of oral cancer.

Results Elevated RNA expression of five metabolic enzymes including GPX2, aldo-keto reductase family 1 members AKR1C1, AKR1C3, and cytochrome monooxygenases (CYP4F11 and CYP4F3) co-occurred in cold tumors across all four smoking-related cancers. These genes have all been linked to negative regulation of arachidonic acid metabolism—a well-established inflammatory pathway—and are also known downstream targets of the redox sensitive Nrf2 transcription factor pathway. In OSCC, LUSC, and LUAD, GPX2 expression was highly correlated with Nrf2 activation signatures, also elevated in cold tumors. In BLCA, however, GPX2 correlated more strongly than Nrf2 signatures with decreased infiltration of multiple leukocyte subtypes. GPX2 inversely correlated with expression of multiple pro-inflammatory cytokines/chemokines and NF-κB activation in cell lines and knockdown of GPX2 led to increased secretion of prostaglandin E2 (PGE2) and interleukin-6. Conversely, GPX2 overexpression led to reduced PGE2 production in a murine OSCC model (MOC1). GPX2

WHAT IS ALREADY KNOWN ON THIS TOPIC

⇒ Immune escape is a hallmark of cancer with multiple mechanisms identified, but the biology behind existence of immunologically cold tumors with little baseline infiltration of leukocytes is poorly understood.

WHAT THIS STUDY ADDS

⇒ We have identified a common set of metabolic genes, including glutathione peroxidase 2, that are at the nexus of arachidonic acid metabolism and neutralization of reactive oxygen species that likely quench intracellular inflammation, shutting down NF-κB signaling and extracellular triggering of host immunity to create immunologically cold tumors despite high mutational neoantigen burdens.

HOW THIS STUDY MIGHT AFFECT RESEARCH, PRACTICE OR POLICY

⇒ We have identified novel potentially targetable enzymes that are likely responsible for a poor host antitumor response, which can be used theranostically to both predict which patient tumors may be poor candidates for immune therapy and to develop new approaches for overcoming treatment failure.

overexpressing MOC1 tumors had a more suppressive tumor immune microenvironment and responded less favorably to anti-cytotoxic T-lymphocytes-associated protein 4 IC therapy in mice.

Conclusion GPX2 overexpression represents a novel potentially targetable effector of immune escape in cold tumors.

BACKGROUND

Antitumor immunity is a critical regulator of tumorigenesis, metastasis and response to conventional chemoradiation regimens across a wide variety of solid tumor histologies. Large scale characterizations of the tumor immune microenvironment (TIME)

across multiple cancer types have consistently revealed at least two kinds of immunosuppressive phenotypes.¹⁻³ In the first category are tumors characterized by abundant infiltration of myeloid derived suppressor cells and regulatory T-cells (T-regs) often present simultaneously with cytotoxic CD8 +lymphocytes, sometimes referred to as immunologically ‘hot’ tumors. In the second category are immunologically ‘cold’ tumors that have minimal baseline leukocyte infiltration. While both scenarios likely contribute to immune tolerance, it is generally believed that hot tumors may respond better to immune checkpoint inhibitors (ICIs) than cold tumors because the former have pre-existing cytotoxic cells potentially capable of recognizing tumor derived neoantigens.⁴⁻⁷ For cold tumors, on the other hand, antitumor immunity is likely abrogated at a much earlier step—making this subtype potentially more challenging to treat with ICIs.

Suppression of the TIME is driven by multiple mechanisms including increased expression of immune checkpoint molecules like programmed death-ligand 1 (PD-L1) on tumors and their cognate programmed cell death protein-1 (PD-1) receptors on lymphocytes, or decreased generation/presentation of mutation-based or expression-based neoantigens by cancer cells.^{8,9} However, there is now increasing evidence that altered metabolic activity inside tumor cells plays a critical role in the recruitment of suppressive cells and the exhaustion of effector cells including CD8 +tumor infiltrating lymphocytes (TILs).¹⁰⁻¹³ A reduction of oxidative phosphorylation and increased production of lactate has broad based suppressive effects on functional immunocytes and has been shown by us and others to generate ‘cold’ tumors depleted of effector cells.^{12,14} Modulation of the kynurenine pathway through differential activation of indoleamine 2,3 dioxygenase has also been shown to cause T-cell anergy, increase proliferation of T-regs and alter the overall balance of the Th1/Th2 response.¹⁵ Just how broadly relevant these phenomena are remains unclear.

We hypothesized that there may be tumor-derived factors other than neoantigen burden driving an immunologically cold phenotype. Using an unbiased bioinformatics approach, we sought to identify candidate genes or pathways that may explain why a substantial subset of human tumors exhibit relatively minimal baseline leukocyte infiltration. Our analysis identified overexpression of glutathione peroxidase 2 (GPX2) to be robustly associated with cold tumors in four different tobacco associated cancers, including squamous cell carcinomas of oral cavity (OCSCC), lung (LUSC), lung adenocarcinomas (LUAD), and bladder cancer (BLCA). Forced overexpression of GPX2 in preclinical models of OCSCC reduced production of inflammatory mediators by tumor cells, generated an aggressive in vivo tumorigenic profile, reduced infiltration of effector immunocytes and reduced effectiveness of cytotoxic T-lymphocytes-associated protein 4 (CTLA-4) inhibition. Collectively, our data indicate that activation of GPX2 represents a previously understudied and

potentially targetable potent driver of a suppressive TIME which could be responsible for limited ICI effectiveness in a significant subset of tumors.

METHODS

Genomic data from TCGA cohorts

Files containing raw RNA-seq by Expectation Maximization (RSEM) count data from the The Cancer Genome Atlas (TCGA) RNA sequencing (RNA-seq) data sets that were previously harmonized computationally and mapped through the Toil open source calling pipeline¹⁶ were downloaded from Amazon web services S3 as previously described.¹⁷ RSEM counts were normalized to get upper quartile Fragments per Kilobase of transcript per Million mapped reads (FPKM) values (FPKM-UQ) with an in-house script in R studio. Essentially, non-protein coding RNAs were removed using annotations from Ensembl, which left 19,345 genes. The 75th percentile RSEM value (excluding zeros) for each TCGA sample was calculated and used to derive FPKM-UQ values. For most statistical analyses, the log₂ (FPKM-UQ +0.01) transformation was applied. TCGA clinical and somatic mutation data were downloaded from cbiportal.¹⁸

Bioinformatic analysis of TIME in TCGA cohorts

Single sample gene set enrichment analysis (ssGSEA) using published gene lists specific for 13 different leukocyte subpopulations was performed using GenePattern software as we described previously,¹² except FPKM normalized RNA-seq data from 20,000 different TCGA samples representing >30 cancer types was used to exclude genes with poor cross-correlation of expression from leukocyte gene lists. The final vetted immune subtype gene lists used for ssGSEA appear in online supplemental table 1. Because only one gene, FOXP3, represented T-regs, the log₂ expression value was used in place of a ssGSEA score in downstream analyses for this leukocyte subtype. To derive immune phenotypes, two-way consensus hierarchical ward’s clustering based on random resampling was performed and optimal cluster numbers chosen using normalized Euclidian distances as we previously described,¹² except a new in-house MATLAB script was employed (see online supplemental methods). To generate final heatmaps, ward’s clustering of the consensus clustering similarity matrices was performed with JMP V.13 software (SAS, Cary, North Carolina, USA).

Cells

Murine oral cancer (MOC) cell line MOC1 was obtained from our collaborator and maintained as previously described.¹⁹ Human head and neck squamous cell (HNSCC) lines, their maintenance, and genomic characterization have also been previously described.²⁰ The viral packaging cell line 293FT was purchased from Thermo Fisher Scientific (Waltham, Massachusetts, USA) and maintained in Dulbecco’s Modified Eagle Medium high glucose containing 10% heat-inactivated fetal bovine

serum, non-essential amino acids (0.1 mM), sodium pyruvate (1 mM), 2 mM L-glutamine, penicillin (100 units/mL) and streptomycin (50 mg/mL).

GPX2 overexpression and knockdown

A mouse complementary DNA (cDNA) encoding GPX2 and the 3' regulatory selenocysteine insertion element was cloned (online supplemental methods) into pLVX-IRES-Puro (online supplemental figure 1), which lacks exogenous tags. For in vitro GPX2 knockdown, human cells were infected with lentiviral pGIPZ-shRNA (Horizon, Waltham, Massachusetts, USA) targeting human GPX2 (V3LHS_368762 with antisense TCCAGGCCACATCTGAGCG) or an empty vector (EV) control. Details regarding infection with lentiviral constructs and isolation of cells with genetically altered GPX2 expression are provided in online supplemental methods.

Western blotting

Cells were lysed in standard radioimmunoprecipitation assay buffer with detergent, and equal amounts of protein diluted in Laemmli SDS loading buffer plus reducing agent were resolved on SDS gels (Invitrogen, Carlsbad, California, USA) before transferring onto polyvinylidene difluoride (PVDF) membrane (Millipore, Temecula, California, USA). Blots were blocked with 5% non-fat milk, washed in TBS-tween, probed with primary antibody, and further washed before developing with horseradish peroxidase (HRPO)-conjugated secondary antibody and the ECL Western detection reagent (Thermo Scientific). Primary antibodies used in this study were as follows: from R&D systems, GPX2 (clone 496010); from Millipore Sigma, β -actin (clone AC-15).

Flow cytometry assessment of TIME in MOC1 tumors overexpressing GPX2

A total of 4×10^6 MOC1/wt, MOC1/pLVX-IRES-puro (EV), or MOC1/pLVX-IRES-puro-GPX2 cells were implanted subcutaneously in the left flank of female C57BL/6J mice (8–10 weeks). Tumors of equal size (minimum five per group) were dissociated in RPMI 1640 (Sigma-Aldrich) containing DNase I (20 U/mL; Sigma-Aldrich), Collagenase I (1 mg/mL; EMD Millipore) and Collagenase IV (250 U/mL; Worthington Biochemical Corporation) using a gentleMACS Dissociator (Miltenyi Biotec, California, USA) and then incubated at 37°C for 30 min to complete digestion. Tumor infiltrating leukocytes were enriched from single cell suspensions using Lymphoprep (STEMCELL Technologies) and analyzed in unblinded fashion using a panel of subset specific markers (online supplemental methods) as we previously described^{11 21} using the gating strategy illustrated in online supplemental figure 2.

In vivo response to immune checkpoint inhibitor anti-cytotoxic T-lymphocytes-associated protein 4

A total of 5×10^6 MOC1/wt, MOC1/pLVX-IRES-puro (EV), or MOC1/pLVX-IRES-puro-GPX2 cells were implanted subcutaneously in the left flank of female C57BL/6J

mice (8–10 weeks). When tumors reached palpable size around day 7 post tumor inoculation, 10 mice were randomized into control or treatment groups. Treatment regimens included a total of three doses each of intraperitoneal injections of either InVivoMAb anti-mouse cytotoxic T-lymphocytes-associated protein 4 (CTLA-4) (clone 9H10; BioXCell; 100 μ g per dose) or InVivoMAb polyclonal Syrian hamster IgG (Isotype Control; BioX-Cell; 100 μ g per dose) once every 5 days. The mice were monitored, in unblinded fashion, two times per week and the tumor growth was measured using calipers. Tumor area (mm^2) was calculated as $A=L \times W$, where L is length and W is width. Mice were sacrificed when their tumors reached maximum allowable burden (above 200 mm^2), if tumors became ulcerated, or animals lost >20% body weight.

Metabolomics analysis

Cells were plated at low density in biologic triplicate and grown exponentially (50%–70% confluence) for 48 hours. Samples were normalized to an internal standard depending on the method used during processing: L-creatinine was used for the Luna method, with a coefficient of variation (CV) of 0.12; L-zeatin (CV of 0.11) was used in the negative method; and L-tryptophan (CV of 0.11) was used for the positive method. After normalization, data were log₂ transformed and then differential expression was determined. Individual metabolite shifts were quantified using linear fold changes compared with the parental cell line. Only metabolites which generated a significant shift (false discovery rate (FDR)<0.1) in two GPX2 overexpression (OE) clones and did not generate a significant shift in the EV expressing cells were further analyzed. Individual metabolite functional analysis and pathway assigned was manually curated using the Human Metabolome Database (V.4.0). Secondary pathway enrichment analysis was performed using MetaboAnalyst (V.5.0).

Other statistical analyses

Differential analysis of gene expression (DEG) between hot and cold tumors was analyzed individually for each TCGA cohort after removing low expression genes (ie, max average log₂ transformed FPKM-UQ for hot or cold groups <2) and performing multiple (two-tailed) t-tests with a Benjamini-Hochberg correction (FDR=0.1) using the Response Screening module in JMP. To find the intersection of up or downregulated genes across all four cohorts a minimum threshold of 1.5-fold change and FDR<0.1 was applied. Multiple t-testing with Benjamini-Hochberg correction or analysis of variance with post-hoc Tukey multiple comparison testing for other data sets were performed with Prism V.8 software (GraphPad, San Diego, California, USA). The Pearson product-moment correlation (ie, Spearman's rho) for gene correlations, along with their FDR-adjusted p values were calculated in JMP. For some analyses, tumor purity values for the TCGA data sets were obtained through the 'TCGA_mastercalls.

abs_tables' file available from the National Cancer Institute (NCI) genomic data commons website.

RESULTS

Identification of immunologically hot and cold tumor samples

TCGA RNA-seq data were used to computationally quantify the relative tumor infiltration levels of leukocytes within solid tumors derived from four different smoking-related cancers, including OCSKC, LUSC, LUAD, and BLCA. Vetted gene lists (online supplemental table 1) defining different leukocyte subpopulations were used to derive ssGSEA scores for each leukocyte subtype in individual tumor samples from all four cancers (online supplemental table 2). For each TCGA cohort, the ssGSEA scores plus RNA values from FOXP1 (ie, Tregs) were then used to perform consensus hierarchical clustering with optimal selection of cluster numbers (ie, online supplemental figure 3) using approaches we previously described.¹² Heatmaps from hierarchical clustering revealed easily identifiable and highly similar patterns of leukocyte infiltration among all four cancer types (figure 1). In each cohort we identified a subset of tumors with high levels of infiltration from nearly every immune subset examined (hot tumors) and at the opposite extreme a substantial number of samples with relatively little infiltration by any of the different leukocytes, which we refer to as 'cold tumors' (figure 1, online supplemental table 2). There were also varying groups of samples with an intermediate phenotype that sometimes separated based on whether cells mediating adaptive immunity or non-adaptive immunity predominated. Tumors with a higher influx of CD8 +T cells or cytotoxic cells in their TIME have frequently been associated with increased survival. Consistent with this observation, we found an increased tendency for improved overall median survival time in patients whose tumors were from cluster 1 (heavy leukocyte infiltration) compared with patients whose tumors had far reduced leukocyte infiltration, across all four tumor types (online supplemental figure 4) although statistical significance was not always reached.

For subsequent analysis, tumor clusters were grouped into the categories of hot, intermediate, or cold tumors based on robustness of pan-leukocyte infiltration (figure 1, online supplemental table 2). As a further validation we performed unbiased DEG expression to confirm higher expression of immune-related genes in hot tumors versus cold tumors. The number of significantly upregulated genes in hot tumors with at least a 1.5-fold average increase ranged from 1710 to 2285 among the four cancers (figure 1, online supplemental tables S3–S6), with 846 commonly upregulated genes (online supplemental table 7). Numerous immune-related genes and known regulators of inflammation, including many not part of gene lists used for ssGSEA, were commonly upregulated across hot tumors. Among these were 20 chemokines/chemokine receptors, 19 interleukins, and

17 human leukocyte antigen (HLA) genes (figure 1, online supplemental table 7). A more intense immune signature has sometimes been linked to a higher mutational tumor burden (ie, increased potential for neoantigens), while other investigators have found the exact opposite relationship depending on the tumor type or study.²² We found no significant differences in the total number of non-synonymous or truncating mutations between hot and cold tumors across all four smoking-related cancers (online supplemental figure 5) that could explain differences in leukocyte infiltration.

Overexpression of enzymes negatively regulating arachidonic acid metabolism dominates the profile of immunologically cold tumors

Next, we focused on genes overexpressed in the cold tumors. Fewer genes were significantly upregulated in the cold tumors (figure 1 and online supplemental tables S3–S6), with LUAD having the least (N=1049) and BLCA the most (N=1628). Only 117 genes were consistently upregulated across cold tumors from all four smoking-related cancers (figure 1, online supplemental table 8) and these were significantly enriched for multiple Gene Ontology (GO) pathways involving metabolism (online supplemental table 9), including oxidoreductase activity (FDR=0.0297). GPX2 showed the highest fold increase in cold tumors when averaged across all four cancer types (9.4-fold), ranging from 6.3-fold to 17.8-fold (figure 1, online supplemental table 8), followed by aldo-keto reductase family 1 member C2 (AKR1C2) which had the second highest average increase in expression (5.5-fold). Other AKR1 members (AKR1C1 and AKR1C3) as well as two cytochrome P450 genes (CYP4F11 and CYP4F3) that encode monooxygenases were also among the 117 common genes upregulated in cold tumors. An anti-inflammatory role of GPX2 is suggested by previous work showing that knockout of the gene in mice increases inflammation and tumorigenesis in an inflammatory model of colon carcinogenesis.²³ Furthermore, depletion of GPX2 expression through siRNA and shRNA reportedly causes increased interleukin (IL)-1 β -mediated activation of NF- κ B, which is associated with increased expression of cyclooxygenase-2 (COX-2) and downstream production of pro-inflammatory prostaglandin E2 (PGE2) in human colon cancer cell lines.^{24–25} All three of the AKR1C family members elevated in cold tumors are capable of reducing PGE2 production by conversion to PGF2a,²⁶ while CYP4F3 and CYP4F11 inactivate leukotriene 4B (a potent inducer of neutrophil chemotaxis) as well as other and inflammatory eicosanoids derived from arachidonic acid metabolism.^{27–28} Although all of these metabolic genes may contribute to the cold tumor phenotype, we focused on GPX2 for further analysis because of its strong link to inflammatory models of colon cancer in mice and previous studies correlating its expression to tumor progression and poor prognosis in multiple human cancers.^{29–31} Levels of GPX2 messenger RNA broadly correlated with the degree of pan-leukocyte

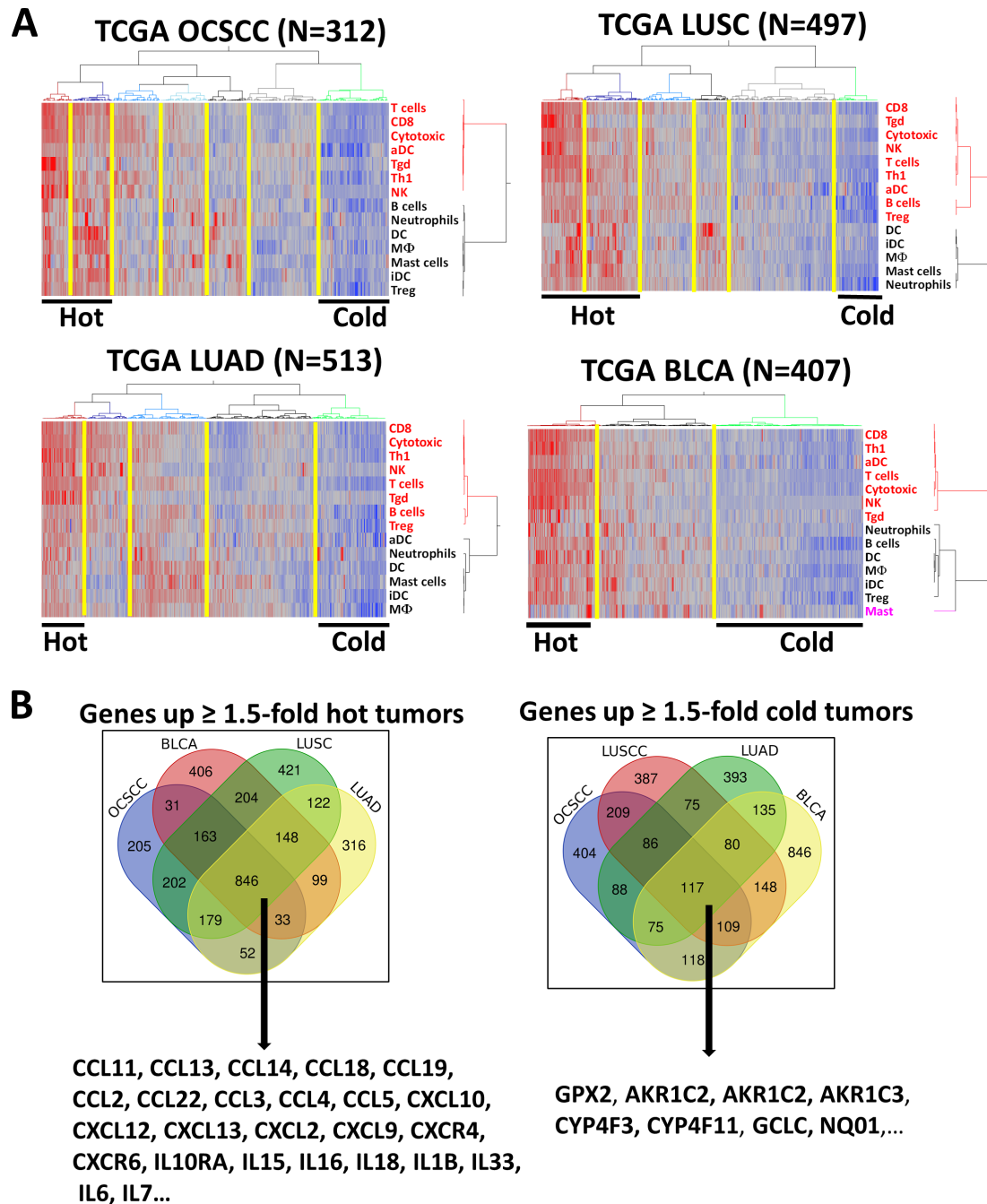


Figure 1 Identifying correlates of tumor immunity in tobacco-related malignancies. (A) Clustering of OCSCE, LUSC, LUAD and BLCA based on single sample gene set enrichment analysis score for 13 leukocyte subtypes (and log2 FOXP3 RNA for regulatory T-cells) identifies subsets of immunologically hot and cold tumors. OCSCE samples included tumors from the oral cavity, oral tongue (excluding base of tongue), floor of mouth, buccal mucosa, alveolar ridge, hard palate, and lip. (B) Differential analysis of genes upregulated in hot and cold tumors across all four tumor histologies. AKR1, aldo-keto reductase family 1; BLCA, bladder cancer; CYP, cytochrome P; GCLC, glutamine-cysteine ligase catalytic subunit; GPX2, glutathione peroxidase 2; LUAD, lung adenocarcinomas; LUSC, lung squamous cell carcinomas; OCSCE, squamous cell carcinomas of oral cavity; NQO1, NAD(P)H quinone dehydrogenase 1.

infiltration among all four smoking-related cancers (online supplemental figure 6A). Because GPX2 has a restricted expression pattern in normal tissues with little RNA found in fibroblasts or leukocytes (online supplemental figure 6B), we examined whether increased tumor purity in cold tumors could account for large increases in GPX2 expression. Applying a conservative approach,

linear GPX2 RNA data were normalized to tumor purity values publicly available for TCGA samples and still found to be significantly elevated in cold tumors (ie, threefold to ninefold increase, FDR<0.01) across all four cancer types (online supplemental figure 6C). Analysis of single cell RNA-seq data (GSE103322) for OCSCE specimens available through the Gene Expression Omnibus (GEO)

data portal confirmed that GPX2 is predominately expressed by tumor cells and not leukocytes or other stromal compartments (online supplemental figure 6D). As further validation, we compared GPX2 expression in hot and cold tumors using data deposited in GEO (ie, GSE81089) from a TCGA-independent Swedish cohort of lung tumors. Using our same ssGSEA and clustering approach for the Swedish cohort (online supplemental figure 7), GPX2 RNA expression was found to be 13.6-fold higher in cold tumors ($p < 0.002$).

GPX2—Nrf2 correlations with TIME

Eleven out of the 117 genes commonly upregulated in cold tumors are known downstream targets of Nrf2 activation (online supplemental table 8),^{32,33} including GPX2, AKR1C family members, CYP4F genes, glutamine-cysteine ligase catalytic subunit (GCLC), and NAD(P)H quinone dehydrogenase 1. Since mutations in the KEAP1/NRF2 pathway have been recently linked to poor leukocyte infiltration,³⁴ we sought to determine whether the correlation between GPX2 and a cold TIME were specific to GPX2 or a byproduct of a more general Nrf2 pathway activation. We first used unbiased gene correlation methods to define a 138 gene signature of Nrf2 activation across all four tumor sites (online supplemental methods and figure 8A). GO analysis (online supplemental table 10) confirmed that this 138 gene signature was enriched for genes involved in processing of oxidative, metabolic, and xenobiotic stress consistent with the function of Nrf2 as a master oxidative stress sensor. The Nrf2-gene set (online supplemental table 11) and ssGSEA scores defined by this signature correlated strongly with Nrf2 pathway mutational status across all four tumor types (online supplemental figure 8B,C). GPX2 expression levels also correlated strongly with Nrf2 ssGSEA scores for OCSCC, LUSC, LUAD and less so for BLCA (online supplemental figure 9). Both GPX2 expression and Nrf2 activation (ie, ssGSEA scores) were significantly correlated with specific immune subsets when examined individually in OCSCC and BLCA (online supplemental figure 10A). However, the significance of correlations was greater for GPX2 than Nrf2 score in OCSCC and both the correlations and the significance levels were stronger for GPX2 than Nrf2 score in BLCA. Collectively, the data suggest that GPX2 was a stronger driver of poor leukocyte infiltration than Nrf2 activation, particularly for BLCA (online supplemental figure 10B).

To determine whether the relationship between GPX2 and Nrf2 is maintained in preclinical models of OCSCC we re-analyzed RNA-seq data from a panel of 65 HNSCC cell lines that we previously published.²⁰ A strong correlation between GPX2 expression and Nrf2 ssGSEA score was found in the HNSCC cell lines (figure 2A, $r = 0.64$; $p < 0.0001$) regardless of HPV and NRF2 mutational status. We also found a strong correlation between GPX2 levels in the cell lines and expression of TP63 (figure 2B), which has previously been described as a transcriptional activator of GPX2.³⁵ To better identify possible regulatory

mechanisms in the cell lines we clustered the expression of GPX2, TP63, Nrf2 score, and GCLC—another downstream target of Nrf2 that strongly correlated with the Nrf2 ssGSEA scores both in cell lines and TCGA samples. Cell lines with the strongest GPX2 levels were found mainly in two clusters (figure 2C), including one with high Nrf2 scores and elevated expression of the Nrf2 target GCLC that was enriched for Nrf2 pathway mutations and another cluster with much lower Nrf2 scores but higher TP63 levels. Additional evidence for Nrf2-independent mechanisms also regulating GPX2 is apparent from clustering TCGA OCSCC specimens by the 117 genes upregulated in cold tumors and their Nrf2 activation scores, which revealed a group of cold tumors with relatively high GPX2 but low Nrf2 activation (online supplemental figure 11). Collectively, the data suggest a dual regulatory mechanism of GPX2 in HNSCC that may involve both stress-induced signaling by Nrf2 and endogenous oncogenic signaling through TP63.

GPX2 reduces inflammatory mediator production in tumors

We hypothesized that GPX2 could negatively regulate intracellular inflammation in tumor cells causing reduced secretion of pro-inflammatory mediators into the surrounding TIME. Conceptually, this would be difficult to validate from in vivo RNA-seq data sets because the mere presence of leukocytes in hot tumors is both a source and cause of inflammatory gene expression. Therefore, we leveraged the RNA data from our panel of HNSCC cell lines,²⁰ representing pure tumor cells, to examine correlations between GPX2 expression and 69 different cytokines/chemokines. Remarkably, 15 different cytokines/chemokines were significantly associated ($FDR < 0.1$) with GPX2 expression in the cell lines (online supplemental table 12 and figure 12A,B). The majority (ie, 14/15) represented pro-inflammatory genes or cytokines that had moderate but significant anti-correlations, supporting our hypothesis that GPX2 negatively regulates inflammation. GPX2 expression in the cell lines was also significantly anti-correlated with NF- κ B activation assessed by ssGSEA (online supplemental figure 12C), consistent with an earlier report that GPX2 reduces NF- κ B activity in colon cancer cell lines. Collectively, the complimentary associations of GPX2 and cytokine/chemokine expression with NF- κ B activation in HNSCC cell lines (online supplemental figure 12D,E), suggest the former may broadly impact inflammatory mediators by attenuating NF- κ B activity.

Among chemokines negatively correlated with GPX2 were molecules strongly chemotactic for neutrophils (CXCL1,2,3) or monocytes, activated CD8 +T cells, NK cells, macrophages and dendritic cells (ie, CCL2 and CCL5). The pro-inflammatory cytokines IL-1A, IL-1B, IL-6, IL-8, IL-32, IL-20, IL-24, IL-23A were also inversely correlated with GPX2 expression. Interestingly, the immune checkpoint PD-L1 was also negatively correlated while IL-17D known to suppress CD8 +T cells was the only cytokine positively correlated with GPX2 expression. To

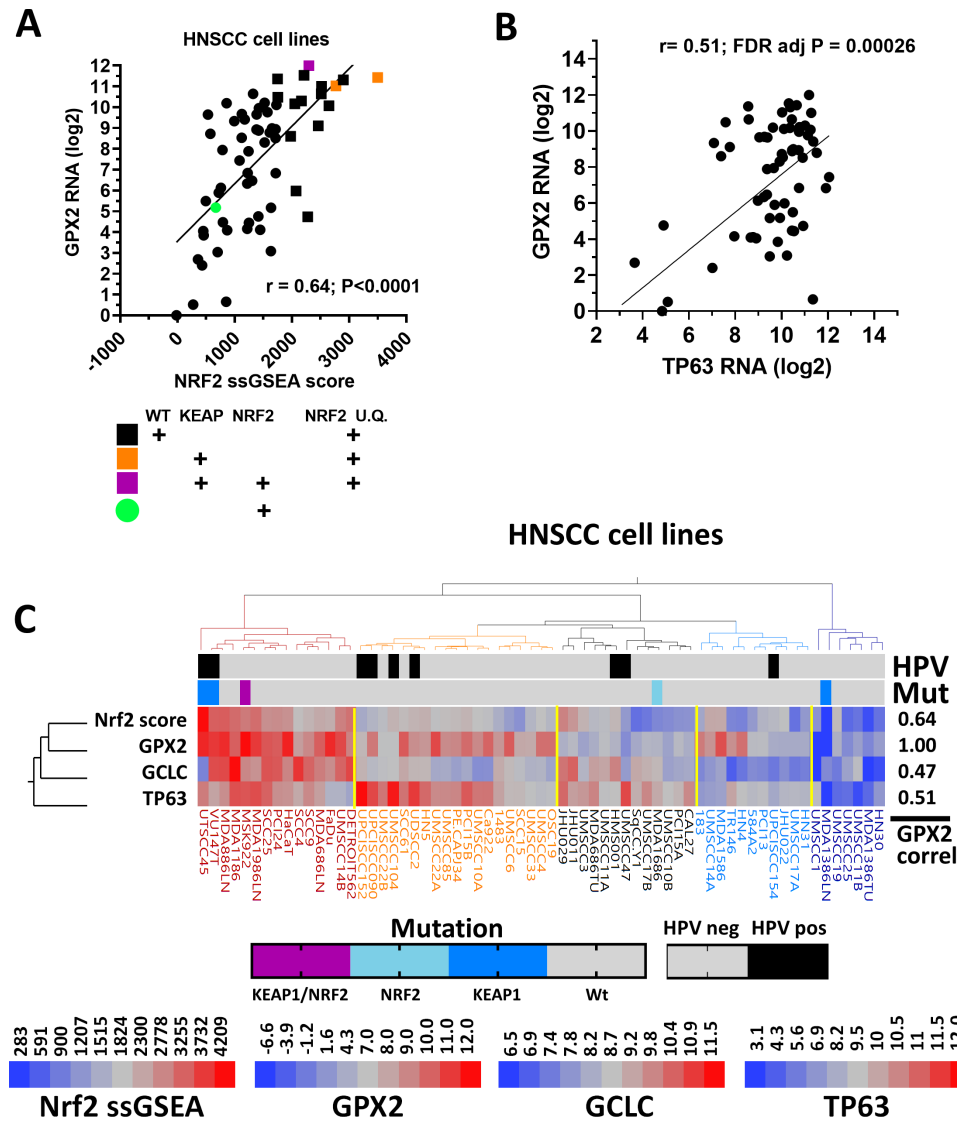


Figure 2 Nrf2-dependent and Nrf2-independent GPX2 regulation in HNSCC tumor lines. (A) Correlation between GPX2 RNA expression and Nrf2 activation (ie, Nrf2 single sample gene set enrichment analysis ssGSEA score) in a large panel of HNSCC cell lines annotated by KEAP1/NRF2 mutational status or relative amount of Nrf2 activation (U.Q., upper quartile). Remaining WT samples with Nrf2 scores below the UQ are represented by black circles. (B) Correlation between GPX2 RNA and TP63 expression in the same panel of HNSCC cell lines. (C) Clustering HNSCC cell lines by Nrf2 activation and expression of GPX2, TP63, or a second downstream Nrf2 target (GCLC) revealed five clusters or patterns of expression. GCLC, glutamine-cysteine ligase catalytic subunit; GPX2, glutathione peroxidase 2; HNSCC, human head and neck squamous cell; ssGSEA, single sample gene set enrichment analysis.

understand better whether *GPX2* expression or Nrf2 activation was a stronger driver of the anti-inflammatory genotype we performed the same analysis using Nrf2 ssGSEA scores derived from the cell lines and compared their anti-correlations and FDR adjusted p-values (online supplemental table 12 and figure 12A). Pro-inflammatory genes were more strongly anti-correlated with GPX2, than with Nrf2 activation scores. Finally, we analyzed correlations with *GPX2* expression in an independent RNA-seq data set from a different smoking-related tumor type consisting of 24 LUSC tumor lines, which are part of the Cell Line Encyclopedia Project. *GPX2* expression was similarly inversely correlated with multiple pro-inflammatory genes in LUSC cell lines (online

supplemental table 13). IL-6, IL-32, along with PD-L1, were all strongly anti-correlated with *GPX2*, along with additional pro-inflammatory genes uniquely significant for the LUSC tumors. The suppressive cytokine IL-17D was also positively correlated with *GPX2* in LUSC cell lines as in HNSCC.

To test the functional link between GPX2 and downstream proinflammatory mediators, we knocked down expression of the former using shRNA in two different human HNSCC cell lines (figure 3A), which expressed high (FaDu) or low (UMSC22A) baseline levels of GPX2 protein. Suppression of *GPX2* with targeting shRNA but not an EV led to a twofold to threefold increase in levels of secreted IL-6 detectable in tumor supernatants collected

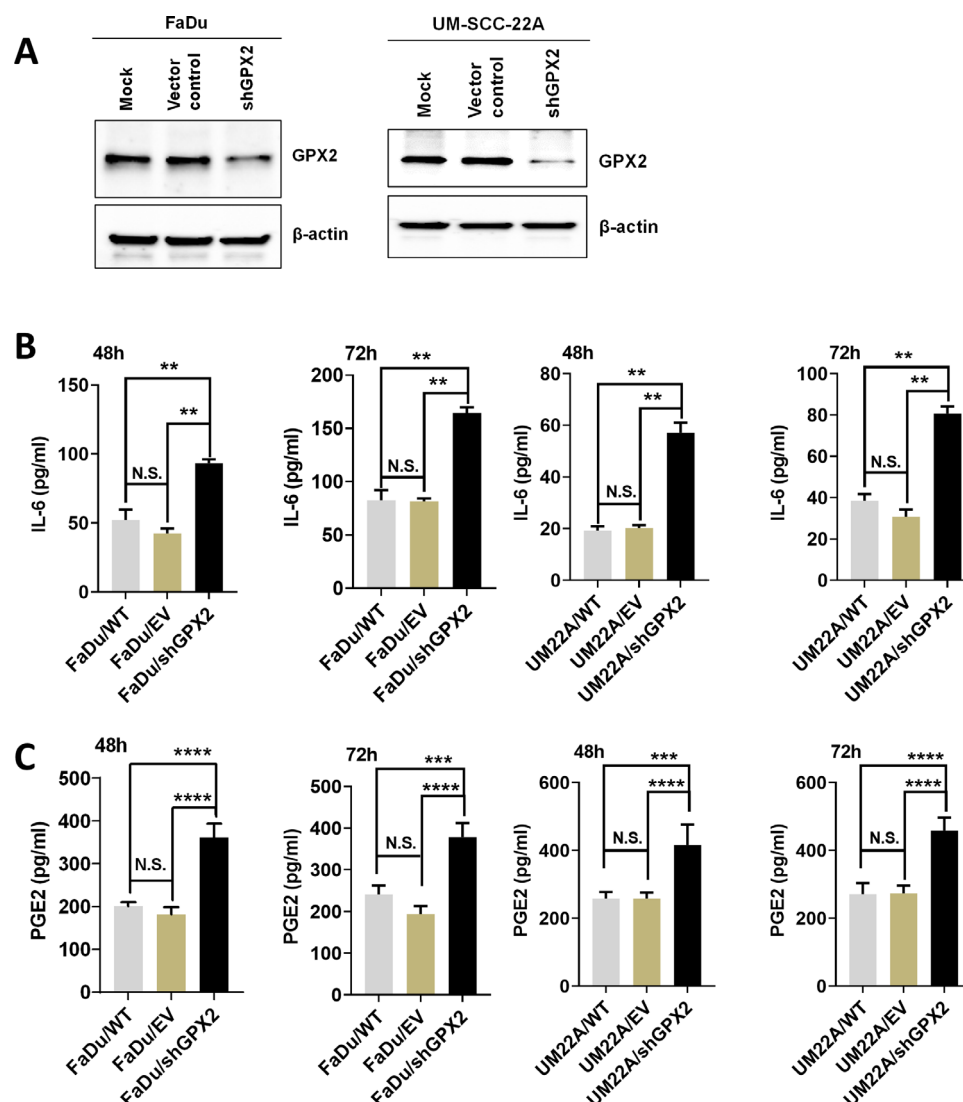


Figure 3 GPX2 suppression increases inflammatory mediator production by HNSCC cell lines. FaDu and UM22A cells were infected with lentiviral constructs containing either empty vector (EV) or shRNA targeting GPX2. (A) Western blot validation of GPX2 protein knockdown 72 hours post infection. Insert demonstrates suppression of GPX2 protein levels at 48 hours post infection. Conditioned media was analyzed at 48 and 72 hours post-infection using ELISA to quantitate levels of secreted IL-6 (B) or PGE2 (C). Soluble mediator levels are normalized to total number of cells. Data are presented as means, with error bars denoting SD. * denotes $p < 0.05$; ** denotes $p < 0.01$; *** denotes $p < 0.001$; **** denotes $p < 0.0001$. IL, interleukin; GPX2, glutathione peroxidase 2; HNSCC, human head and neck squamous cell; PGE2, prostaglandin E2.

at 48 hours and 72 hours following knockdown in both FaDu and UM22A (adj $p < 0.01$, [figure 3B](#)). GPX2 knockdown also resulted in increased secretion (1.6–1.8-fold) of proinflammatory PGE2 from both cell lines at each time point (adj $p < 0.001$, [figure 3C](#)). Conversely, we overexpressed mouse GPX2 cDNA in MOC1, a preclinical murine model of oral carcinoma previously shown to have relatively high leukocyte infiltration when grown in vivo.¹⁹ Increased GPX2 protein expression by western blotting and a doubling of intracellular peroxidase activity in overexpressing clones were confirmed ([figure 4A,B](#)). Consistent with the human cell line data, *overexpressing GPX2* drove levels of PGE2 in the opposite direction resulting in more than a twofold drop ([figure 4C](#)). IL-6 levels in parental cells were not detectable in either control MOC1 or after GPX2 OE.

GPX2 regulates cellular metabolism

To test whether *GPX2* OE shifted cellular metabolism in a manner consistent with its previously described function of a stress response enzyme we performed steady state metabolomics analysis. *GPX2* OE cells demonstrated a dramatic decrease in carnitine derivatives (online supplemental table 14) consistent with suppression of fatty acid beta oxidation, decreased levels of oxidized glutathione, glucose and disruption of multiple metabolic regulators of oxidative stress including taurine (−7.7, −4.5 linear decreases, respectively for each of 2 *GPX2* OE clones, compared with MOC parental line), betaine (−3.7, −3.4, respectively), octulose-8-phosphate (1.8, 2.1, respectively) and sialic acid (3.6, 3.5, respectively). Several metabolites with known immunomodulatory activity were significantly increased in *GPX2* OE cells

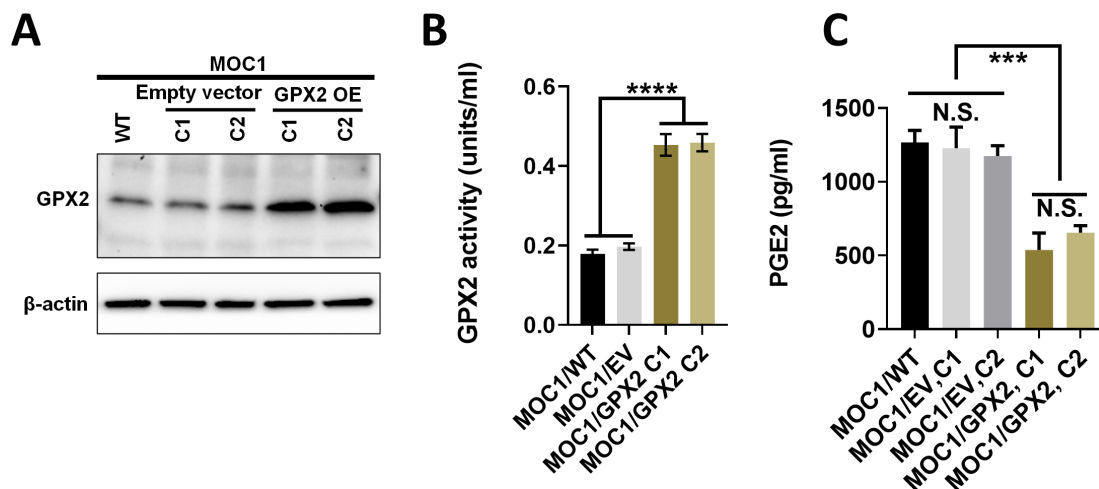


Figure 4 GPX2 overexpression suppresses PGE2 production in murine HNSCC cell lines. (A) Western blot validation of murine GPX2 OE in MOC1 cells. Clones 1 and 2 (C1 and C2) were derived from polyclonal stable cell lines selected with puromycin following infection with either an empty vector or one with the mouse GPX2 complementary DNA insert. (B) GPX2 OE doubled the intracellular GPX2 enzymatic activity measured in MOC1 cell lines. (C) PGE2 secretion was reduced by twofold in MOC1 clones with GPX2 OE. Stable overexpression of either empty vector or GPX2 containing constructs were accomplished in the MOC1 murine cell line (insert denotes Western blot). Individual clones (C) were established for both the empty vector control (EV) and for the GPX2 overexpressing constructs, and reduced PGE2 secretion. Data are presented as means, with error bars denoting SD. *** denotes $p < 0.001$; **** denotes $p < 0.0001$. GPX2, glutathione peroxidase 2; HNSCC, human head and neck squamous cell; PGE2, MOC, murine oral cancer; OE, overexpression; PGE2, prostaglandin E2.

including 3- hydroxykynurenine (2.1, 2.5, respectively), phenylalanine (2.3, 2.2, respectively) and phosphocholine (2.7, 2.1, respectively). These metabolic shifts were paralleled by changes in expression of metabolic genes measured through RNA-seq. GO enrichment analysis of genes significantly upregulated in both *GPX2* OE clones compared with the EV and parental cell lines mapped to multiple metabolic pathways (online supplemental table 15) including the following pathways with enrichment ≥ 2 fold and $FDR < 0.0001$: peptide metabolic process, translation and cellular nitrogen compound biosynthetic process. Genes downregulated in *GPX2* OE cells mapped to pathways (enrichment ≥ 2 fold and $FDR < 0.05$) including: double-stranded break repair via non-homologous end joining, inter-strand cross-link repair, glycoprotein biosynthetic process, phospholipid biosynthetic process, cellular response to stress and cellular response to DNA damage stimulus. The following genes mapped (Reactome) to selenocysteine synthesis were significantly upregulated in both OE clones: RPL13, FAU, RPS29, RPL18A, RPL12, RPL5, RPL9, RPL29 and UBA52.

GPX2 drives shifts in TIME and impacts ICI response

In three independent experiments (figure 5, online supplemental figure 13), tumors grown subcutaneously from MOC1 overexpressing *GPX2* in immunocompetent C57BL/6J mice had significantly fewer overall percentage of T cells and decreased percentages of M1-macrophages, with reduced levels of CD8 +T cells significant in 2/3 experiments and trending towards significance in a third replicate experiment. Conversely, MOC1 *GPX2* OE tumors had significantly increased levels (ie, almost

double) of infiltrating granulocyte-myeloid derived suppressor cells (G-MDSC) compared with MOC1 EV or parental MOC1 in both experiments. For most leukocyte subsets, no statistically significant differences were found between the MOC1 parental and MOC1 EV tumors, but the average percentages were sometimes intermediate in value between MOC1 and MOC1 *GPX2* OE, suggesting that the presence of viral vector sequences could have weakly impacted the TIME.

We tested whether the altered TIME in MOC1 *GPX2* OE tumors, characterized by reduced T-cell infiltration and increased presence of G-MDSC, associated with *GPX2* OE would impair antitumor efficacy of immune checkpoint therapy. In two independent experiments, single agent blockade of CTLA-4 was less effective at blocking tumor progression in MOC1 *GPX2* OE tumors compared with MOC1 EV control. In a pooled analysis from the two experiments (figure 6), just 11/20 (ie, 55%) mice with MOC1 *GPX2* OE tumors responded compared with 19/20 (ie, 95%) of mice harboring MOC1 EV tumors ($p < 0.009$). This was accompanied by a highly significant reduction of survival time ($p < 0.005$) following anti-CTLA-4 treatment in mice harboring tumors overexpressing *GPX2*. Interestingly, the anti-CTLA-4 response observed for MOC1 EV tumors was more robust than previously reported for MOC1 parental tumors,³⁶ raising the possibility that the presence of viral vector sequences does impart some added immunogenicity that was reversed by *GPX2* OE. Nevertheless, MOC1 EV tumors grew in 100% of control (isoform) treated mice, although the *GPX2* OE tumors were comparatively more aggressive in the absence of anti-CTLA-4 ($p < 0.0001$, figure 6C, D and E). Our preclinical

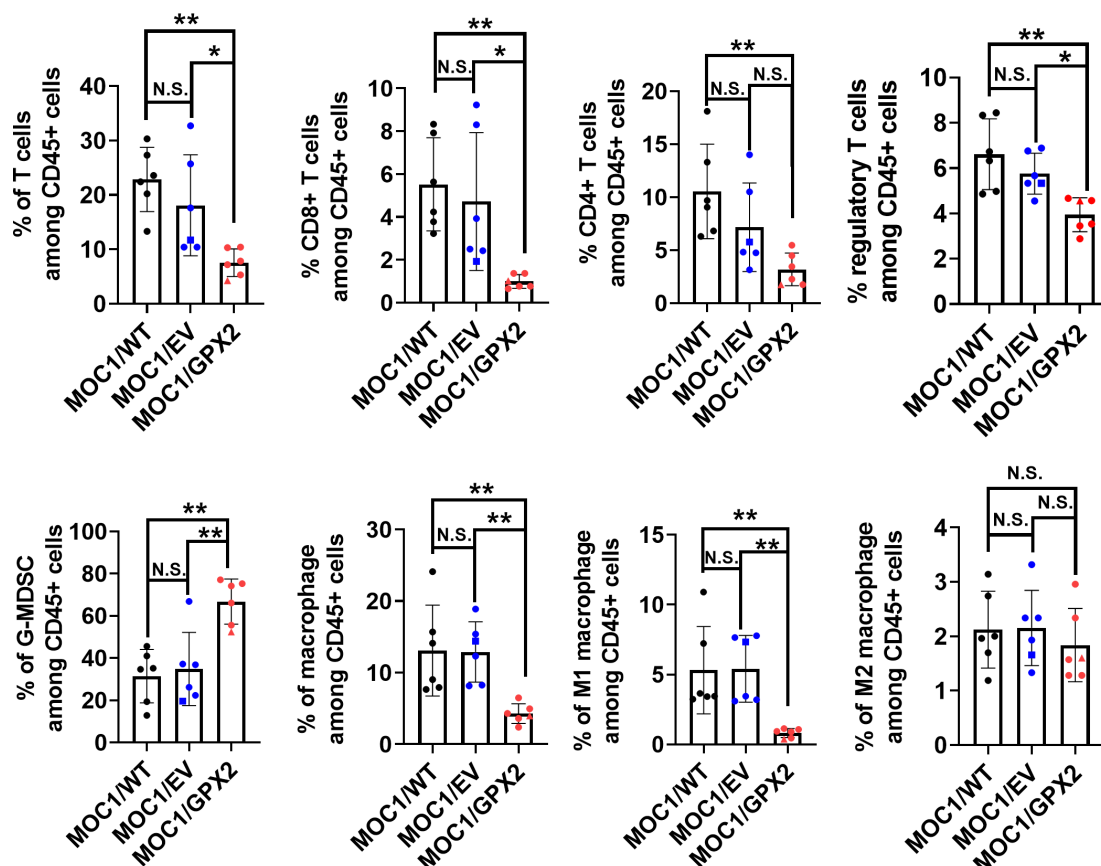


Figure 5 GPX2 drives tumor immune microenvironment shifts. MOC1 tumors (N=6) generated from parental (wild-type (WT)), empty vector (EV) or GPX2 containing constructs were harvested at the same size and analyzed via flow cytometry. All data are presented as means, with error bars denoting SD and individual tumor volumes showed using individual circles. * denotes $p < 0.05$; ** denotes $p < 0.01$; *** denotes $p < 0.001$, **** denotes $p < 0.0001$, after performing a Tukey multiple comparison test and further adjusting p values with a Benjamini-Hochberg correction ($FDR < 0.1$) to control the family wise error within a cell line group due to multiple testing across different leukocyte subpopulations. The arcsine transformation was applied to percentages prior to statistical analysis. EV, empty vector; G-MDSC, granulocyte-myeloid derived suppressor cells; GPX2, glutathione peroxidase 2; MOC, murine oral cancer.

data suggest that patients with higher GPX2 expression may be less responsive to ICIs. To that end, we examined the relationship between GPX2 expression in TCGA samples (OCSCC and BLCA) and their tumor inflammatory score (TIS) derived using an 18 gene pan-cancer immune signature that robustly predicts ICI response across multiple cancers.³⁷ We found highly significant inverse correlations ($p < 5 \times 10^{-11}$) between levels of GPX2 RNA and the TIS (online supplemental figure 14). In BLCA the anti-correlation with TIS was much greater for GPX2 than Nrf2, supporting that the immunophenotype was more strongly driven by GPX2.

GPX2 expression is a poor prognostic marker in OCSCC

Consistent with an immunologically cold phenotype, high levels of GPX2 RNA were strongly correlated with worse survival in the OCSCC TCGA cohort. Patients with OCSCC with high GPX2 expression in their tumors had a median survival time of just 32.4 months compared with 71.2 months for patients with low GPX2 expression ($p = 0.0092$, figure 7), which represents a greater than twofold difference. High Nrf2 activation was similarly

associated with a substantial reduction in survival time as well ($p = 0.039$), although it was statistically not as significant as GPX2 expression.

Pan-cancer analysis of GPX2 expression in cold tumors

The link between GPX2 expression and cold tumors was examined in 23 other cancer types and in the two other main subsites of head and neck cancer: squamous cell carcinomas of the oropharynx (OPSCC) and larynx/hypopharynx (LHSCC). Although differences in GPX2 expression did not correlate with the TIME in OPSCC (online supplemental figure 15A,B), a majority of these tumors harbored high-risk oncogenic human papilloma virus-16 which was itself strongly associated with hot tumors ($p = 0.02$). Analysis of LHSCC identified significant OE of arginase 2 in cold tumors (online supplemental figure 15C), rather than GPX2, suggesting these tumors utilize an alternative metabolic mechanism for immunosuppression. We examined the correlation between GPX2 RNA expression and leukocyte infiltration (based on ssGSEA scores) across an additional 23 cancer types (online supplemental figure 16A) to identify tumors

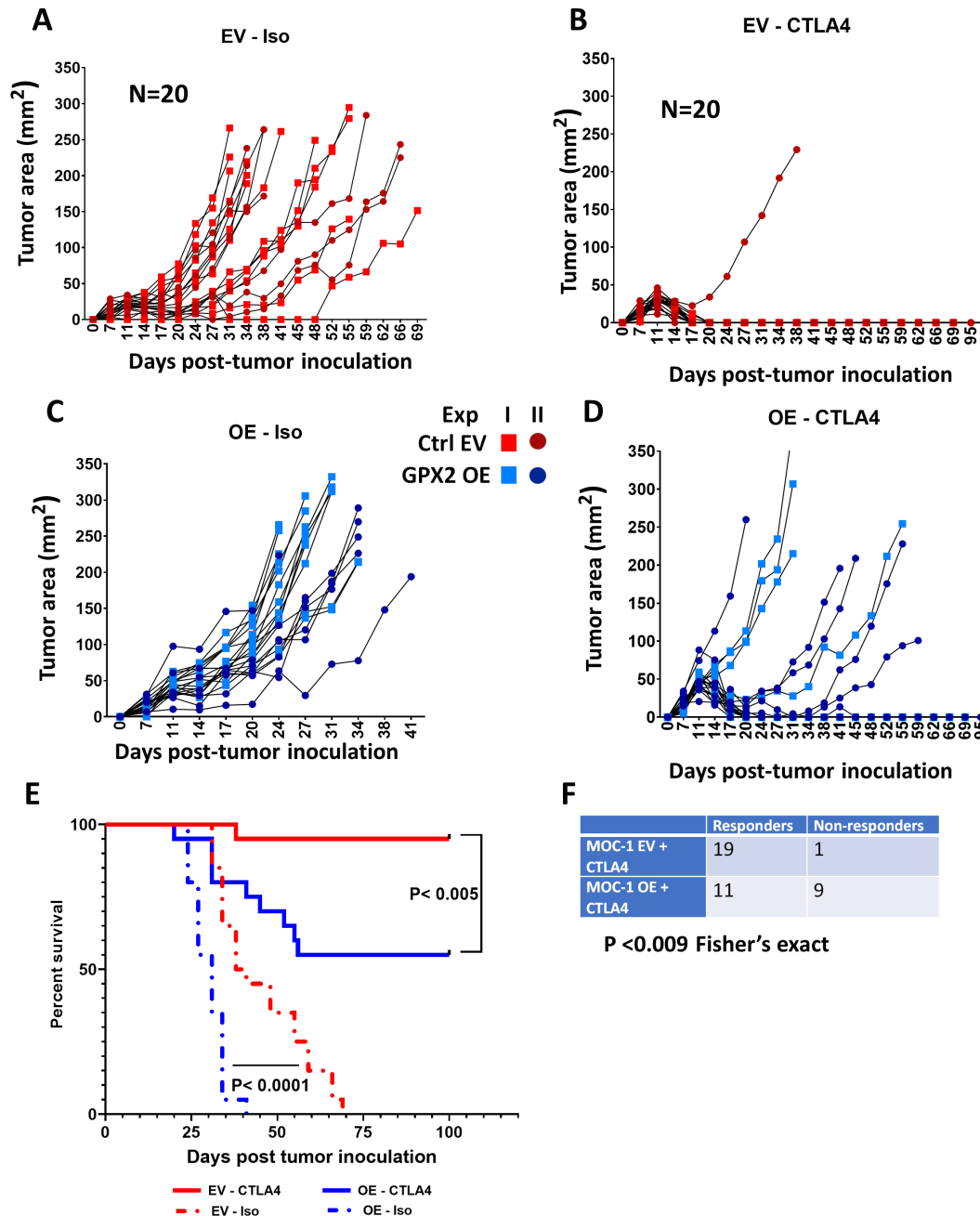


Figure 6 GPX2 reduces ICI effectiveness. Pooled analysis of two independent experiments in which MOC1 tumors with EV (A,B) or GPX2 OE (C,D) were established and allowed to grow in the flank model. Established tumors underwent treatment with three total injections of either isotype (iso) control (A,C) or anti-CTLA4 antibody (B,D). Tumor measurements are presented as individual symbols (square=experiment 1; circle=experiment 2) and each tumor is represented as an individual curve for the growth panels (red=EV; blue=OE). Each treatment group had 20 tumors (exp I=12; exp II=8) (E) Survival is denoted using Kaplan-Meier curves as an aggregate of the treatment groups. (F) Table of responders and non-responders following anti-CTLA4 antibody treatment. CTLA-4, cytotoxic T-lymphocytes-associated protein 4; EV, empty vector; GPX2, glutathione peroxidase 2; ICI, immune checkpoint inhibitor; MOC, murine oral cancer; OE, overexpression.

that behaved similarly to smoking-related cancers. Levels of GPX2 RNA (after adjusting for tumor purity) were significantly elevated in cold tumors compared with hot tumors in colon and rectal adenocarcinomas (online supplemental figure 16B,C), but the opposite relationship was found for a few cancers where GPX2 expression levels were comparatively lower, such as prostate adenocarcinoma (online supplemental figure 16A,D). In other

cancers, such as cholangiocarcinoma, *moderate* inverse correlations with GPX2 and leukocyte infiltration did not persist after adjusting for tumor purity (data not shown).

DISCUSSION

For nearly a decade we have been both dazzled by the amazing effectiveness of ICIs in the context of previously

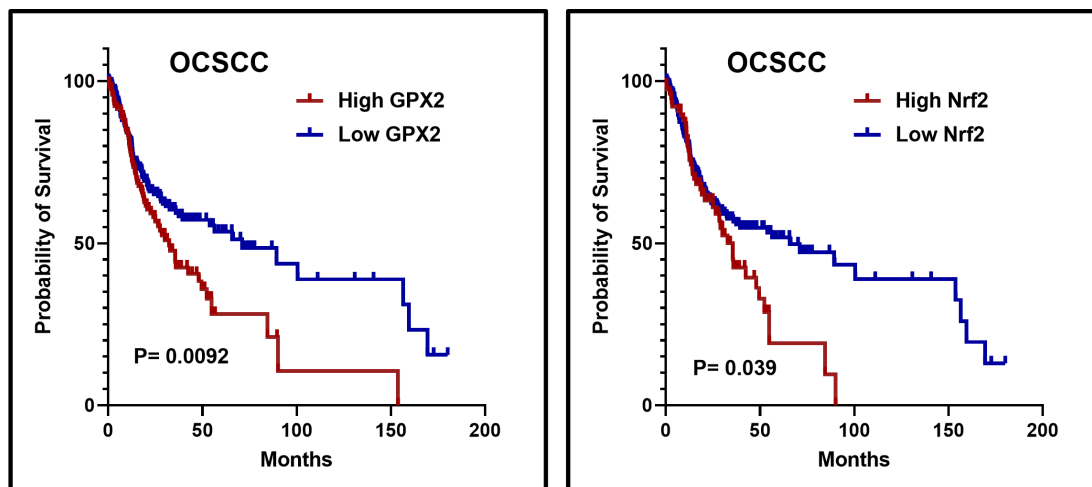


Figure 7 Both GPX2 and Nrf2 activation correlate with poor survival in OSCC. Recursive partitioning was used to dichotomize patient tumors as either high or low for GPX2 RNA expression (left) or Nrf2 activation (right). Kaplan-Meier curves of overall survival were analyzed for statistical significance with a log-rank test. HRs for tumors expressing high GPX2 and Nrf2 activation are 1.593 and 1.452, respectively. GPX2, glutathione peroxidase 2; OSCC, squamous cell carcinomas of oral cavity.

untreatable solid tumors (eg, metastatic melanoma) and disappointed when these effects fail to materialize in less immunogenic tumors such as OSCC.³⁸ We have come to realize that tumor cells are able to modulate their immunological phenotype much like they can internal signaling networks (eg, protein phosphorylation cascades). ICI effectiveness is critically linked to the presence and activation of TILs.⁴ Similarly, CAR-T therapies require that lymphocytes infiltrate and maintain activation inside of solid tumors. In both cases, if tumor cells can either prevent infiltration or suppress activation once TILs have reached the tumor, the therapy will be ineffective. Our data indicate that GPX2 activation likely does both by preventing secretion of inflammatory mediators. In particular, the potential of GPX2 to blunt very early stages of tumor inflammation by suppressing pro-inflammatory chemokines and cytokines sets it apart mechanistically from the more widely studied processes of immune escape tied to expression of immune checkpoint molecules like PD-L1.

GPXs are ubiquitously expressed selenium-dependent enzymes that protect cells against oxidative damage by reducing hydrogen peroxide and a wide range of organic peroxides using glutathione as a primary source of reducing potential. GPX2, one of the eight members of the GPX family (GPX1–8), is also called gastrointestinal GPX because of its relatively high expression in the gastrointestinal tract. Because of their antioxidative activity, several isoforms of GPXs have been investigated in association with inflammation and cancer.³⁹ GPX2 in particular has been shown to play an important role in the progression of malignant tumors with GPX2 OE detected in liver,⁴⁰ colorectal,^{41 42} breast,⁴³ lung⁴⁴ and castration-resistant prostate tumors.³¹ In the context of tumorigenesis, a preventive role of GPX2 has been proposed because mice, in which both GPX1 and GPX2 had been knocked out, progressively developed ileocolitis

and subsequently intestinal cancer, especially when not raised under specific pathogen-free conditions.⁴⁵

As with all other regulators of oxidative stress, expression of GPX2 likely generates a push-stop phenomenon whereby sufficient oxidative stress is allowed for tumor cell proliferation, growth and metastasis but excess levels are kept in check to prevent premature tumor cell death. It has been described as a tobacco response gene in both lung and oral cavity epithelium.^{44 46} GPX2 and AKR1Cs both have both been linked to development of resistance to chemotherapy and radiation in multiple tumor models including HNSCC and LUAD.^{47 48} We previously showed that HNSCC cell lines adapted to oxidative stress rewire metabolism⁴⁹ in a manner similar to that observed when GPX2 is exogenously overexpressed in the current manuscript.

Adaptation to oxidative stress, whether induced by genotoxic agents (eg, cisplatin or radiation), inflammatory cells (eg, neutrophils) or tobacco exposure likely triggers fundamental metabolic reprogramming which involves, at least in part hyperactivation of GPX2. What we have now shown mechanistically is that when this happens, there are important immunologic consequences. OE of GPX2 in an immune competent HNSCC model not only reduced effector TIL levels, but also increased levels of suppressive immunocytes including G-MDSC. Future experiments with additional mouse tumor lines where GPX2 is manipulated are needed to dissect how existing genetic background and intracellular tumor signaling may be influenced by GPX2 to alter the inflammatory secretome. One possible way GPX2 may shift the TIME is by attenuating NF- κ B mediated inflammation as suggested by prior reports in which GPX2 knockdown in human colon cancer cells reduced NF- κ B activation, downstream expression of COX-2, and increased PGE2 synthesis²⁵ NF- κ B is stimulated by intracellular reactive oxygen species,⁵⁰ and GPX2 is thought

to inhibit NF- κ B indirectly by reducing ROS.²⁵ In support of this model, we found a plethora of pro-inflammatory cytokines and chemokines whose expression inversely correlated with GPX2 levels in a large panel of established human HNSCC lines. Many of these negatively correlating cytokines, including IL-1, IL-6, IL-8, CCL2, CCL5, CXCL1, and CXCL2 are all downstream targets of NF- κ B activation. We have confirmed in vitro that GPX2 knockdown reduced IL-6 production in two different human HNSCC cell lines tested. Our model suggests that GPX2 OE in tumors limits inflammation by dampening initial recruitment and subsequent activation of effectors from both innate and adaptive immunity. This model could account for why GPX2 is linked to the cold TIME in a subset of tumor types. Possibly, NF- κ B-induced inflammation may be more relevant to smoking-related cancers and colorectal tumors. In cancer types with a comparatively lower range of GPX2 expression, such as prostate adenocarcinoma, neutralization of intracellular ROS may be incomplete and thus elevated GPX2 would become a surrogate for unresolved intracellular inflammation that leads to increased leukocyte recruitment.

While this explanation is inherently simple, it may in fact be incomplete as several of the metabolites which increase following constitutive GPX2 OE have been shown to directly regulate immunocyte activity (eg, taurine, kynurenine derivatives). Basic shifts in energy and biomass balance have been shown by us and others to correlate with shifts in TIME at least in part to differential balance of oxidative phosphorylation and lactate producing glycolytic activity.¹² Data generated here suggest that additional metabolic shifts may further complicate this metabolic link between tumor cells and immunocytes. Furthermore, a number of other enzymes including the AKR1 family members and cytochrome monooxygenases were upregulated in parallel with GPX2 in cold tumors. These enzymes can directly regulate the arachidonic acid cascade, altering production of pro-inflammatory prostaglandins and leukotrienes, complementing the ability of GPX2 to inhibit PGE2 expression via COX2 downregulation.

Because GPX2 has been functionally linked to chemotherapy resistance⁴⁷ we hypothesize that treatment failure to first line genotoxic therapies will select for patient tumors with a metabolic phenotype less likely to respond to subsequent ICIs. Two future directions remain our current focus. First, development of small molecule GPX2 inhibitors is critical to further preclinical validation of this immunomodulatory model in combination with clinically available ICIs. Second, expansion of functional genomic experiments to include other Nrf2-regulated metabolic enzymes upregulated in cold tumors will allow us to dissect their individual and collective contributions to modulating the TIME in preclinical models of HNSCC, BLCA, and LUAD. Such experiments will help clarify how best to clinically target these novel putative metabolic drivers of immunological escape in smoking-related cancers.

Author affiliations

¹Bobby R. Alford Department of Otolaryngology - Head and Neck Surgery, Baylor College of Medicine, Houston, Texas, USA

²Department of Head and Neck Surgery, Division of Surgery, University of Texas MD Anderson Cancer Center, Houston, Texas, USA

³Department of Molecular and Cellular Biology, Baylor College of Medicine, Houston, Texas, USA

⁴Advanced Technology Core, Dan Duncan Cancer Center, Baylor College of Medicine, Houston, TX, USA

⁵Human Genome Sequencing Center, Baylor College of Medicine, Houston, Texas, USA

⁶Department of Pathology and Immunology, Baylor College of Medicine, Houston, TX, USA

⁷Undergraduate School of Engineering, Cornell University, Ithaca, New York, USA

⁸The University of Texas at Austin School of Biological Sciences, Austin, Texas, USA

⁹ENT Section, Operative Care Line, Michael E. DeBakey Veterans Affairs Medical Center, Houston, TX, USA

¹⁰Center for Translational Research on Inflammatory Diseases, Michael E. DeBakey Veterans Affairs Medical Center, Houston, TX, USA

Contributors MAK: Performed experiments and analyzed data. RV: Performed experiments and analyzed data. DD: Performed experiments and analyzed data. NRP: Performed experiments and analyzed data. VP: Performed experiments and analyzed data. MFC: Wrote code, downloaded and analyzed data. DAW: Provided intellectual input and edited manuscript. WD: Provided intellectual input and edited manuscript. AIF: Developed and algorithms and analyzed data. SK: Performed experiments and analyzed data. AS: Provided intellectual input and edited manuscript. VS: Performed experiments, analyzed data, intellectual input, and writing. MJF: Project conception, data analysis, writing, intellectual input, supervision. MJF as the guarantor for the current manuscript accepts full responsibility for the finished work and/or the conduct of the study, had access to the entire dataset, and controlled the decision to publish.

Funding VCS is supported by a Career Development Award from the Veterans Administration Clinical Science Research and Development division (1IK2CX001953) and an American Cancer Society Research Scholar Grant. This work was supported by the National Institute of Dental and Craniofacial Research through R03DE028858 (VCS) and the Sid W. Richardson Foundation (MJF&VCS). This project was also supported by the Cytometry and Cell Sorting Core at Baylor College of Medicine with funding from the CPRIT Core Facility Support Award (CPRIT-RP180672), the NIH (P30 CA125123 and S10 RR024574) and the expert assistance of Joel M. Sederstrom.

Disclaimer The contents of this manuscript do not represent the views of the U.S. Department of Veterans Affairs or the United States Government.

Competing interests None declared.

Patient consent for publication Not applicable.

Ethics approval Human subjects research restricted to public databases and therefore exempt; all mouse studies performed with institutionally approved IACUC protocols.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data are available in a public, open access repository. Files containing recalculated RNA-seq FPKM-UQ values derived from publically available TCGA datasets corresponding to the TCGA cohorts used in this study will be made available upon individual request.

Supplemental material This content has been supplied by the author(s). It has not been vetted by BMJ Publishing Group Limited (BMJ) and may not have been peer-reviewed. Any opinions or recommendations discussed are solely those of the author(s) and are not endorsed by BMJ. BMJ disclaims all liability and responsibility arising from any reliance placed on the content. Where the content includes any translated material, BMJ does not warrant the accuracy and reliability of the translations (including but not limited to local regulations, clinical guidelines, terminology, drug names and drug dosages), and is not responsible for any error and/or omissions arising from translation and adaptation or otherwise.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is

properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See <http://creativecommons.org/licenses/by-nc/4.0/>.

ORCID iDs

Vlad C Sandulache <http://orcid.org/0000-0002-9205-385X>

Mitchell J Frederick <http://orcid.org/0000-0002-0932-8809>

REFERENCES

- Thorsson V, Gibbs DL, Brown SD, et al. The immune landscape of cancer. *Immunity* 2019;51:411–2.
- Anichini A, Tassi E, Grazia G, et al. The non-small cell lung cancer immune landscape: emerging complexity, prognostic relevance and prospective significance in the context of immunotherapy. *Cancer Immunol Immunother* 2018;67:1011–22.
- Mandal R, Şenbabaoglu Y, Desrichard A, et al. The head and neck cancer immune landscape and its immunotherapeutic implications. *JCI Insight* 2016;1:e89829.
- Liu Y-T, Sun Z-J. Turning cold tumors into hot tumors by improving T-cell infiltration. *Theranostics* 2021;11:5365–86.
- Taefehshok S, Parhizkar A, Hayati S, et al. Cancer immunotherapy: challenges and limitations. *Pathol Res Pract* 2022;229:153723.
- Oguejiofor K, Galletta-Williams H, Dovedi SJ, et al. Distinct patterns of infiltrating CD8+ T cells in HPV+ and CD68 macrophages in HPV-opharyngeal squamous cell carcinomas are associated with better clinical outcome but PD-L1 expression is not prognostic. *Oncotarget* 2017;8:14416–27.
- Cristescu R, Mogg R, Ayers M, et al. Pan-tumor genomic biomarkers for PD-1 checkpoint blockade-based immunotherapy. *Science* 2018;362:3593. doi:10.1126/science.aar3593
- Suresh S, O'Donnell KA. Translational control of immune evasion in cancer. *Trends Cancer* 2021;7:580–2.
- Goodman AM, Kato S, Bazhenova L, et al. Tumor mutational burden as an independent predictor of response to immunotherapy in diverse cancers. *Mol Cancer Ther* 2017;16:2598–608.
- Krupar R, Hautmann MG, Pathak RR, et al. Immunometabolic determinants of chemoradiotherapy response and survival in head and neck squamous cell carcinoma. *Am J Pathol* 2018;188:72–83.
- Veeramachaneni R, Yu W, Newton JM, et al. Metformin generates profound alterations in systemic and tumor immunity with associated antitumor effects. *J Immunother Cancer* 2021;9:e002773.
- Frederick M, Skinner HD, Kazi SA, et al. High expression of oxidative phosphorylation genes predicts improved survival in squamous cell carcinomas of the head and neck. *Sci Rep* 2020;10:6380.
- Ippolito L, Morandi A, Giannoni E, et al. Lactate: a metabolic driver in the tumour landscape. *Trends Biochem Sci* 2019;44:153–66.
- Romero-Garcia S, Moreno-Altamirano MMB, Prado-Garcia H, et al. Lactate contribution to the tumor microenvironment: mechanisms, effects on immune cells and therapeutic relevance. *Front Immunol* 2016;7:52.
- Johnson TS, McGaha T, Munn DH. Chemo-immunotherapy: role of indoleamine 2,3-dioxygenase in defining immunogenic versus tolerogenic cell death in the tumor microenvironment. *Adv Exp Med Biol* 2017;1036:91–104.
- Vivian J, Rao AA, Nothhaft FA, et al. Toil enables reproducible, open source, big biomedical data analyses. *Nat Biotechnol* 2017;35:314–6.
- Vivian J, Rao A, Nothhaft FA. Rapid and efficient analysis of 20,000 RNA-seq samples with Toil. *bioRxiv* 2016.
- Cerami E, Gao J, Dogrusoz U, et al. The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Discov* 2012;2:401–4.
- Judd NP, Allen CT, Winkler AE, et al. Comparative analysis of tumor-infiltrating lymphocytes in a syngeneic mouse model of oral cancer. *Otolaryngol Head Neck Surg* 2012;147:493–500.
- Kalu NN, Mazumdar T, Peng S, et al. Genomic characterization of human papillomavirus-positive and -negative human squamous cell cancer cell lines. *Oncotarget* 2017;8:86369–83.
- Newton JM, Hanoteau A, Sikora AG. Enrichment and characterization of the tumor immune and non-immune microenvironments in established subcutaneous murine tumors. *J Vis Exp* 2018:57685. doi:10.3791/57685
- Wang X, Li M. Correlate tumor mutation burden with immune signatures in human cancers. *BMC Immunol* 2019;20:4.
- Kreihl S, Loewinger M, Florian S, et al. Glutathione peroxidase-2 and selenium decreased inflammation and tumors in a mouse model of inflammation-associated carcinogenesis whereas sulfuraphane effects differed with selenium supply. *Carcinogenesis* 2012;33:620–8.
- Banning A, Kipp A, Schmitmeier S, et al. Glutathione peroxidase 2 inhibits cyclooxygenase-2-mediated migration and invasion of HT-29 adenocarcinoma cells but supports their growth as tumors in nude mice. *Cancer Res* 2008;68:9746–53.
- Koeberle SC, Gollwitzer A, Laoukili J, et al. Distinct and overlapping functions of glutathione peroxidases 1 and 2 in limiting NF-κB-driven inflammation through redox-active mechanisms. *Redox Biol* 2020;28:101388.
- Dozier BL, Watanabe K, Duffy DM. Two pathways for prostaglandin F2 alpha synthesis by the primate periovulatory follicle. *Reproduction* 2008;136:53–63.
- Johnson AL, Edson KZ, Totah RA, et al. Cytochrome P450 ω-Hydroxylases in inflammation and cancer. *Adv Pharmacol* 2015;74:223–62.
- Johnson AL, Edson KZ, Totah RA. Chapter eight - cytochrome P450 ω-hydroxylases in inflammation and cancer. In: Hardwick JP, ed. *Advances in pharmacology*. Academic Press, 2015: 223–62.
- Liu T, Kan X-F, Ma C, et al. GPX2 overexpression indicates poor prognosis in patients with hepatocellular carcinoma. *Tumour Biol* 2017;39:101042831770041.
- Liu K, Jin M, Xiao L, et al. Distinct prognostic values of mRNA expression of glutathione peroxidases in non-small cell lung cancer. *Cancer Manag Res* 2018;10:2997–3005.
- Naiki T, Naiki-Ito A, Asamoto M, et al. GPX2 overexpression is involved in cell proliferation and prognosis of castration-resistant prostate cancer. *Carcinogenesis* 2014;35:1962–7.
- Banning A, Deubel S, Kluth D, et al. The Gl-GPx gene is a target for Nr2f2. *Mol Cell Biol* 2005;25:4914–23.
- Kim JW, Botvinnik OB, Abudayyeh O, et al. Characterizing genomic alterations in cancer by complementary functional associations. *Nat Biotechnol* 2016;34:539–46.
- Cai M-C, Chen M, Ma P, et al. Clinicopathological, microenvironmental and genetic determinants of molecular subtypes in KEAP1/NRF2-mutant lung cancer. *Int J Cancer* 2019;144:788–801.
- Yan W, Chen X. GPX2, a direct target of p63, inhibits oxidative stress-induced apoptosis in a p53-dependent manner. *J Biol Chem* 2006;281:7856–62.
- Clavijo PE, Moore EC, Chen J, et al. Resistance to CTLA-4 checkpoint inhibition reversed through selective elimination of granulocytic myeloid cells. *Oncotarget* 2017;8:55804–20.
- Ayers M, Lunceford J, Nebozhyn M, et al. IFN-γ-related mRNA profile predicts clinical response to PD-1 blockade. *J Clin Invest* 2017;127:2930–40.
- Ferris RL, Blumenschein G, Fayette J, et al. Nivolumab for recurrent squamous-cell carcinoma of the head and neck. *N Engl J Med* 2016;375:1856–67.
- Brigelius-Flohé R, Kipp A. Glutathione peroxidases in different stages of carcinogenesis. *Biochim Biophys Acta* 2009;1790:1555–68.
- Suzuki S, Pitchakarn P, Ogawa K, et al. Expression of glutathione peroxidase 2 is associated with not only early hepatocarcinogenesis but also late stage metastasis. *Toxicology* 2013;311:115–23.
- Chiu S-T, Hsieh F-J, Chen S-W, et al. Clinicopathologic correlation of up-regulated genes identified using cDNA microarray and real-time reverse transcription-PCR in human colorectal cancer. *Cancer Epidemiol Biomarkers Prev* 2005;14:437–43.
- Murawaki Y, Tsuchiya H, Kanbe T, et al. Aberrant expression of selenoproteins in the progression of colorectal cancer. *Cancer Lett* 2008;259:218–30.
- Naiki-Ito A, Asamoto M, Hokaiwado N, et al. Gpx2 is an overexpressed gene in rat breast cancers induced by three different chemical carcinogens. *Cancer Res* 2007;67:11353–8.
- Woenckhaus M, Klein-Hitpass L, Grepmeier U, et al. Smoking and cancer-related gene expression in bronchial epithelium and non-small-cell lung cancers. *J Pathol* 2006;210:192–204.
- Chu F-F, Esworthy RS, Chu PG, et al. Bacteria-induced intestinal cancer in mice with disrupted GPx1 and GPx2 genes. *Cancer Res* 2004;64:962–8.
- Na H-K, Kim M, Chang S-S, et al. Tobacco smoking-response genes in blood and buccal cells. *Toxicol Lett* 2015;232:429–37.
- Du H, Chen B, Jiao N-L, et al. Elevated glutathione peroxidase 2 expression promotes cisplatin resistance in lung adenocarcinoma. *Oxid Med Cell Longev* 2020;2020:7370157.
- Chang W-M, Chang Y-C, Yang Y-C, et al. AKR1C1 controls cisplatin-resistance in head and neck squamous cell carcinoma through cross-talk with the Stat1/3 signaling pathway. *J Exp Clin Cancer Res* 2019;38:245.
- Yu W, Chen Y, Putluri N, et al. Acquisition of cisplatin resistance shifts head and neck squamous cell carcinoma metabolism toward neutralization of oxidative stress. *Cancers* 2020;12:12061670. doi:10.3390/cancers12061670
- Gloire G, Legrand-Poels S, Piette J. NF-kappaB activation by reactive oxygen species: fifteen years later. *Biochem Pharmacol* 2006;72:1493–505.

Supplementary Methods:

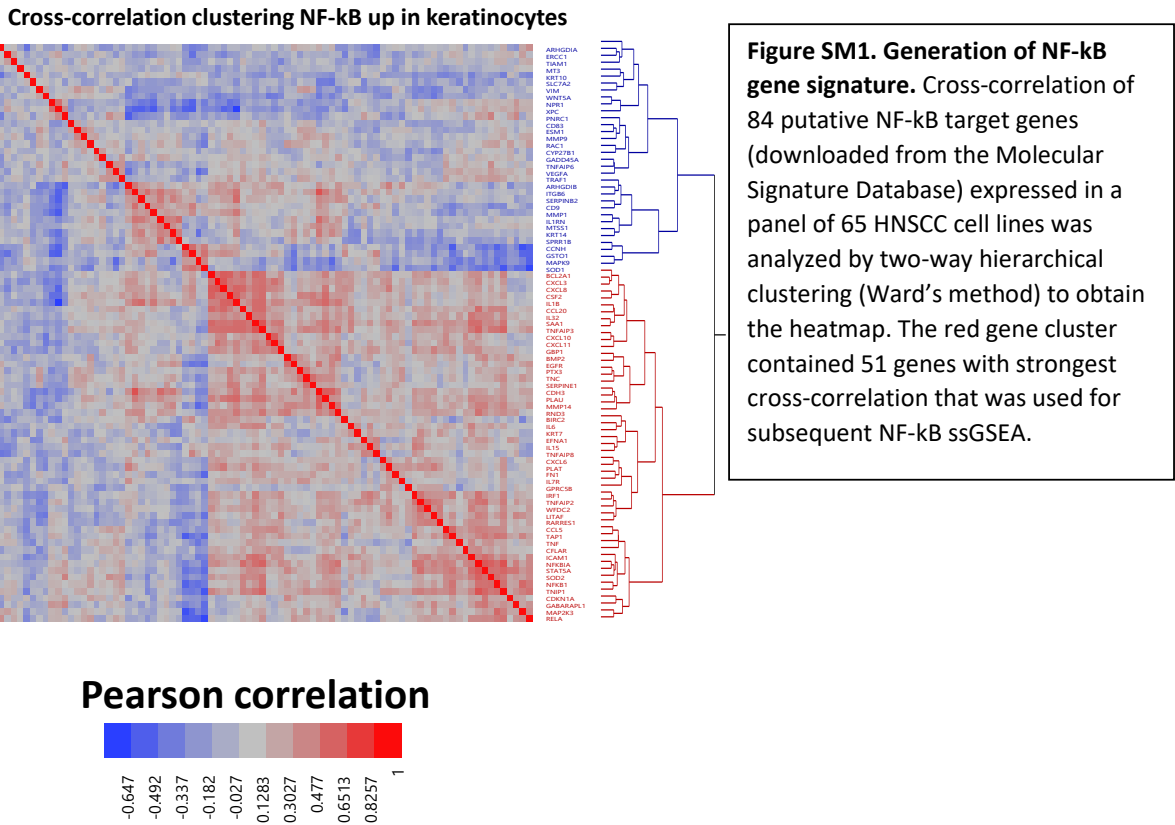
Cloning and viral infection of cells. In-Fusion Cloning (TaKaRa, Osaka, Japan) was used to construct a murine full length expression cDNA construct containing the endogenous 3' UTR corresponding to the regulatory selenocysteine insertion element (SECIS), which directs incorporation of selenocysteine (SEC) into the stop codon positioned at amino acid 40 near the N-terminus. Essentially, RNA was extracted from murine esophageal tissue (C57BL/6J) and cDNA generated using the superscript II reverse transcription kit (Thermo Fisher, Waltham, MA). cDNA was used as a template for PCR amplification of the full length GPX2 cDNA plus the SECIS using primers with additional overlapping sequence from the lentiviral destination vector pLVX-IRES-Puro (Addgene, Watertown, MA). The GPX2-SECIS amplicon was cloned into pLVX-IRES-Puro digested with XhoI / BamHI to generate pLVX-IRES-mGFPX2 (Supplementary Figure 1), which lacks fluorescent or exogenous tags, and the sequence confirmed by Sanger Sequencing. For *in vitro* GPX2 knockdown experiments, human cells were infected with lentiviral pGIPZ-shRNA (Horizon, Waltham, MA) targeting human GPX2 (V3LHS_368762 with antisense TCCAGGCCACATCTGAGCG) or an empty vector (EV) control.

Lentiviral vectors with insert or devoid of insert (negative controls) were packaged in 293FT by co-transfecting them with helper vectors pMD2.G and psPAX2 (Addgene). Supernatants containing virus were collected and concentrated using Lenti-X Concentrators (Takara Bio, USA) before infecting cells by spinoculation in the presence of polybrene (6-10 µg/ml). To generate stable cell lines, MOC1 infected cells were selected with puromycin (2 µg/ml) initially as polyclonal populations and single cells subsequently seeded into 96 well plates using a BD FACS Arial cell sorter to derive long term clones subsequently screened for GPX2 protein levels by Western blot analysis. For GPX2 KD, human cells infected with shRNA were sorted for EGFP expression 48 h post-infection and allowed to recover for 2 days before setting up experiments to collect supernatants for ELISA or cell lysates for Western blot analyses. Levels of prostaglandin E2 (PGE2) or interleukin-6 (IL-6) in conditioned media were measured by human or mouse ELISA kits (Cayman chemicals, Ann Arbor, MI) and normalized to total cell number.

Immune Profiling by flow cytometry. All *in vivo* experiments were performed with approval of the Institutional Animal Care and Use Committee (IACUC) at Baylor College of Medicine (BCM) or MD Anderson and followed established protocols. Female C57BL/6J mice (8–10 weeks of age) were purchased from the Jackson Laboratory and housed under specific pathogen-free conditions. Mice were randomly divided into treatment groups and subcutaneously (s.c.) inoculated with 3×10^6 MOC1/wt, MOC1/pLVX-IRES-puro, or 2.5×10^6 MOC1/pLVX-IRES-puro-GPX2 cells in the right flank. Tumor area (mm²) was calculated as $L \times W$, where L is Length and W is Width. To profile tumor immune cell infiltration, MOC1/wt, MOC1/pLVX-IRES-puro, and MOC1/pLVX-IRES-puro-GPX2 tumors were harvested at equal sizes (~0.8 cm width), and analyzed by flow cytometry using the approach we previously described^{1,2} with a slight modification. Briefly, tumors were dissociated in RPMI 1640 (Sigma-Aldrich) containing DNase I (20 U/ml; Sigma-Aldrich), Collagenase I (1mg/ml; EMD Millipore) and Collagenase IV (250 U/ml; Worthington Biochemical Corporation) using a gentleMACS Dissociator (Miltenyi Biotec, CA) and then incubated at 37°C for 30 minutes to complete digestion. Tumor infiltrating leukocytes were enriched from single cell suspensions using Lymphoprep™ (STEMCELL Technologies). Cells were first blocked with anti-mouse CD16/CD32 Fc block (BD Biosciences) and stained using fluorescently tagged antibodies from either the lymphocyte or myeloid panels (Supplementary Methods). DAPI was added to stained leukocytes to exclude dead cells and subsets analyzed with the gating strategy outlined in Supplementary Figure 2. For intracellular staining, cells were fixed and permeabilized with Intracellular Fixation and Permeabilization

Buffer Set (eBioscience) prior to the addition of anti-INOS. Data were acquired on a LSRII and LSRFortessa (BD Biosciences) flow cytometers, for myeloid and T cell panels respectively, and analyzed using FlowJo v10 software (FlowJo, LLC).

Generation of Nrf2 and NF-κB activation gene signatures. We previously described a method for vetting published gene signature lists to identify the most robust targets specific for a given tissue.³ Essentially, the two by two Pearson correlation values for co-expression among all possible gene pairs are calculated using a cohort of samples with similar histology or the cohort under study and the matrix of cross-correlations is analyzed by two-way hierarchical clustering to produce a heatmap with co-expression modules appearing along the diagonal. Modules are then further evaluated for greatest average cross-correlations or biological significance that matches known functional properties of the pathway being defined to identify the most robust gene list. To derive Nrf2 and NF-κB signatures, we downloaded published lists of 469 genes (NFE2L2.V2) and 84 genes (HINATA NFκB Targets Keratinocytes Up) from the Molecular Signature Database.⁴ Cross-correlation of the 469 Nrf2 gene list was examined in pooled TCGA samples (i.e., Supplementary Figure S7) and the 84 NF-κB gene list in a panel of 65 established human HNSCC cell lines (Supplementary Methods Figure SM1) to identify the most robust downstream targets.



Consensus Clustering—We used a modification of the resampling-based method published by Monti et al.⁵ to perform hierarchical two-way agglomerative (Ward’s linkage) consensus clustering with an in-house Matlab script listed below and publicly available at <https://github.com/aif33/Hierarchical-two-way-agglomerative-consensus-clustering>

To run the program, 5 Matlab files below (names in bold) are created and saved to the same folder containing input data files. Users run the ClusteringScript.m file and input 3 parameters as “File name.xlsx”, “maximum number of clusters to try (i.e., 8)”, and “FirstDimension , Second Dimension or an empty string (””) referring to orientation of Z-scores to be calculated. If input values are already Z transformed, then input the empty string. The program always clusters whichever variable is listed vertically by whatever feature is listed horizontally. For example, to cluster samples based on subset scores or gene expression values the input file matrix would have each row representing a different sample and each column a different feature. If data are not Z-scaled the user would input “FirstDimension “ to calculate Z vertically. To reverse and cluster genes or leukocyte subsets by samples, simply transpose the input matrix and input “SecondDimension” so Z -scores will now be calculated horizontally to give equivalent values. The expected architecture for input files is shown by example below. The additional Matlab files contain helper methods and classes to run Clusterinscript.m. The program is hard coded to use 400 re-samplings but that can be changed within the code according to user preference. The program uses an 80% proportion for random re-sampling and applies the identical set of 400 re-samplings across all “N” clusters to facilitate choice of optimal N, which can be chosen from the output graphical plot of Normalized Euclidian Distance (NED) verses number of clusters by selecting a local minimum. The NED is a measure of error that represent the Euclidean distance between a perfect transformed similarity matrix and the calculated transformed similarity matrix at each value of N as we previously described.³ Thus optimal choice of clusters proceeds independently for the first and second dimension analyses when performing two-way clustering.

FirstDimension input-Z scores calculated vertically

samples	aDC	Bcells	CD8	Cytotoxic DC	iDC	
TCGA-4P-AA8J	3962.401	3861.641	1518.665	1517.294	-2148.8	2101.675
TCGA-BA-4074	3569.409	-1372.16	-3636.74	-2287.41	-1142.81	-1614.04
TCGA-BA-4075	-178.691	-2850.88	-4107.13	-2803.09	807.9017	1071.645
TCGA-BA-5149	6265.759	1085.697	-963.355	-910.477	-2614.58	-936.648
TCGA-BA-5151	6960.077	2522.59	-2192.68	-1244.51	486.7138	291.1622
TCGA-BA-5152	8171.622	3530.965	1547.833	4269.614	2501.658	3267.812
TCGA-BA-5556	8832.716	4422.375	4229.137	5707.508	-432.351	2391.95
TCGA-BA-5557	4666.945	-887.898	588.2251	755.4258	2897.251	2814.198
TCGA-BA-5558	6710.69	1855.682	1482.251	2441.415	-1206.88	537.9367

SeondDimension input- Z scores calculated horizontally

samples	TCGA-4P-AA8J	TCGA-BA-4074	TCGA-BA-4075	TCGA-BA-5149	TCGA-BA-5151	TCGA-BA-5152	TCGA-BA-5556	TCGA-BA-5557	TCGA-BA-5558
aDC	3962.401	3569.409	-178.691	6265.759	6960.077	8171.622	8832.716	4666.945	6710.690
Bcells	3861.641	-1372.162	-2850.881	1085.697	2522.590	3530.965	4422.375	-887.898	1855.682
CD8	1518.665	-3636.743	-4107.126	-963.355	-2192.683	1547.833	4229.137	588.225	1482.251
Cytotoxic	1517.294	-2287.413	-2803.089	-910.477	-1244.508	4269.614	5707.508	755.426	2441.415
DC	-2148.800	-1142.811	807.902	-2614.576	486.714	2501.658	-432.351	2897.251	-1206.876
iDC	2101.675	-1614.045	1071.645	-936.648	291.162	3267.812	2391.950	2814.198	537.937

1. ClusteringScript.m

```

clust = clusterSampling("NewBLCAsamples.xlsx", 8, "FirstDimension");
%possible inputs -> "FirstDimension", "SecondDimension", ""
maxGroups = size(clust,2);

x = zeros(1,maxGroups);
y = zeros(1,maxGroups);

for i = 1:maxGroups
    x(1,i) = i;
    y(1,i) = clust(i).EuclideanDistance;
end

plot(x,y);

```

2. clusterSampling.m

```

function clustSampling = clusterSampling(dataFileName, numberOfGroups,
zScoreDimensionType)

%{

inputs and outputs:

dataFileName --> file name to find the data with sample names in the
left most
column, features we are measuring in the top most row, and values in
the intersections

    feature1  feature2  feature3... featureN
sample1      a         b         c         d
sample2      e         f         g         h
.            .         .         .         .
.            .         .         .         .
.            .         .         .         .
sampleN      i         j         k         l

numberOfGroups -> number of groups to be used in clustering

clustSampling will be a list of a class 'ClusterInfo',
where each instance of 'ClusterInfo' contained in the list represents
instance number N, for 1 <= N <= numberOfGroups.

an instance n of 'ClusterInfo' contains:
(a) percent similarity matrix,
(b) transformed and sorted percent similarity matrix
(c) euclidean distance between matrix of ones and (b)

%}

clustSampling = clusterN.empty;

```

```
[NUM,TXT,RAW] = xlsread(dataFileName);

%eliminate first column and first row labels. then we are going to
randomly
%sample 80 percent of the data and calculate z scores of this randomly
%selected 80 percent.

numberOfSamples = size(NUM,1);

numOfResamplings = 400;

allRks = zeros(numOfResamplings, numberOfSamples);

% generate indexes of which samples we selected randomly with
% 80 % of samples selected
% r(k) == 0 ? picked sample k : did not pick sample k
% where k is the index of the sample in the vector of sample names from
% our data file
% get 400 pages of z scores --> see numOfResamplings

numberOfRandomRows = round(numberOfSamples * .8);

zScoresSampledPages = zeros(numberOfRandomRows, size(NUM,2),
numOfResamplings);

for cnt = 1:numOfResamplings

    k = randperm(numberOfSamples, numberOfRandomRows);
    r = true(1,numberOfSamples);
    r(k) = false;

    allRks(cnt,:) = r;

    numSampled = double.empty;

    for i = 1:numberOfSamples
        if r(i) == 0 %if r is 0, we picked that sample
            numSampled(end+1,:) = NUM(i,:);
        end
    end

    if zScoreDimensionType == "FirstDimension"

        zScoresSampled = zscore(numSampled,0,1);
        %z score of randomly sampled data in first dimension

        zScoresSampledPages(:,cnt) = zScoresSampled(:,:);

    elseif zScoreDimensionType == "SecondDimension"

        zScoresSampled = zscore(numSampled,0,2);
        %z score of randomly sampled data in second dimension

        zScoresSampledPages(:,cnt) = zScoresSampled(:,:);
    else
```

```
        zScoresSampled = numSampled; %input is already z scored

        zScoresSampledPages(:, :, cnt) = zScoresSampled(:, :);
    end

end

for N = 1:numberOfGroups

    numeratorMatrix = zeros(numberOfSamples, numberOfSamples);

    denominatorMatrix = zeros(numberOfSamples, numberOfSamples);

    for cnt = 1:numOfResamplings

        r = allRks(cnt, :);

        zScoresSampled = zScoresSampledPages(:, :, cnt);

        clusterMatrixOfSampled = clusterMatrix(zScoresSampled, N);

        %we are going to take the randomly sampled groups and create a
        %one column
        %matrix with the rest of the groups we did not sample and assign
        "fake"
        %group numbers to the groups that we did not sample

        fakeGroupNumber = 10000;

        indexOfSampled = 1;

        clusterMatrixAll = zeros(i, 1);

        clusterMatrixAllForDenominator = zeros(i, 1);

        for k = 1:numberOfSamples

            if r(k) == 0
                clusterMatrixAll(k) =
clusterMatrixOfSampled(indexOfSampled);
                clusterMatrixAllForDenominator(k) = 1;
                indexOfSampled = indexOfSampled + 1;
            else
                clusterMatrixAll(k) = fakeGroupNumber;
                clusterMatrixAllForDenominator(k) = fakeGroupNumber;
                fakeGroupNumber = fakeGroupNumber + 1;
            end

        end

        similarityMatrixNumerator = similarityMatrix(clusterMatrixAll);

        similarityMatrixDenominator =
similarityMatrix(clusterMatrixAllForDenominator);

        numeratorMatrix = numeratorMatrix + similarityMatrixNumerator;
```



```
        denominatorMatrix = denominatorMatrix +
similarityMatrixDenominator;

    end

    sampleNames = TXT(2:end,1);

    percentSimilarityMatrix = numeratorMatrix./denominatorMatrix;

    clusterSimilarity = clusterMatrix(percentSimilarityMatrix,N);

    %now we will compute transformed percent similarity matrix

    sampleIndexAndGroups = zeros(numberOfSamples,2);

    sampleIndexAndGroups(:,1) = 1:numberOfSamples;

    sampleIndexAndGroups(:,2) = clusterSimilarity;

    sampleIndexSortedByGroups = sortrows(sampleIndexAndGroups,2);

    mappedSampleIntersection =
containers.Map('KeyType','char','ValueType','double');

    percentSimilaritySortedAndT = cell(numberOfSamples +1,numberOfSamples
+1);

    for a = 1:numberOfSamples
        %a is our row index

        groupNumberRow = clusterSimilarity(a);

        rowName = string(sampleNames(a));

        percentSimilaritySortedAndT(a+1,1) =
sampleNames(sampleIndexSortedByGroups(a,1));

        for b = 1:numberOfSamples
            %b is our column index

            colName = string(sampleNames(b));

            groupNumberCol = clusterSimilarity(b);

            if (groupNumberRow == groupNumberCol)
                value = percentSimilarityMatrix(a,b);
            else
                value = 1 - percentSimilarityMatrix(a,b);
            end

            mappedSampleIntersection(rowName + '_' + colName) = value;

            percentSimilaritySortedAndT(1,b+1) =
sampleNames(sampleIndexSortedByGroups(b,1));

        end
    end
```

```

end

for a = 1:numberOfSamples
    rowName = string(percentSimilaritySortedAndT(a+1,1));

    for b = 1:numberOfSamples
        colName = string(percentSimilaritySortedAndT(1,b+1));

        percentSimilaritySortedAndT(a+1,b+1) =
num2cell(mappedSampleIntersection(rowName + '_' + colName));

    end

end

percentSimilarityValuesSortedAndT =
cell2mat(percentSimilaritySortedAndT(2:numberOfSamples + 1, 2:numberOfSamples
+ 1));

%now we will compute the euclidean distance

matrixOfOnes = ones(numberOfSamples);

x = matrixOfOnes - percentSimilarityValuesSortedAndT;

y = x.^2;

sumFinal = sum(y,'All');

finalValue = (sqrt(sumFinal) / numberOfSamples) * 100;

clustSampling(N) = clusterN(percentSimilarityMatrix,
percentSimilaritySortedAndT, finalValue);

end

end

```

3. clusterN.m

```

classdef clusterN
    properties
        PercentSimilarityMatrix
        PercentSimilarityMatrixSortedAndT
        EuclideanDistance
    end

    methods
        function obj =
clusterN(percentSimilarity,percentSimilaritySortedAndT, euclideanDistance)
            obj.PercentSimilarityMatrix = percentSimilarity;
            obj.PercentSimilarityMatrixSortedAndT =
percentSimilaritySortedAndT;

```

```
        obj.EuclideanDistance = euclideanDistance;
    end
end
end
```

4. similarityMatrix.m

```
function D = similarityMatrix(clusterMatrixAll)

N = size(clusterMatrixAll,1);
D = zeros(N,N);
for i=1:N
    for j=1:i
        if (clusterMatrixAll(i) == clusterMatrixAll(j))
            D(i,j) = 1;
        else
            D(i,j) = 0;
        end
        D(j,i) = D(i,j);% D is the similarity matrix
    end
end
```

5. clusterMatrix.m

```
function clustMatrix = clusterMatrix(inputMatrix, numberOfGroups)

%get cluster matrix from N groups, defined by numberOfGroups

linkageMatrix = linkage(inputMatrix, 'ward');
clustMatrix = cluster(linkageMatrix,'MaxClust',numberOfGroups);
end
```

Immune Profiling- Single cell suspensions prepared as described in the main methods were divided into two tubes and stained with fluorescently tagged antibodies to either a lymphocyte or myeloid panel.

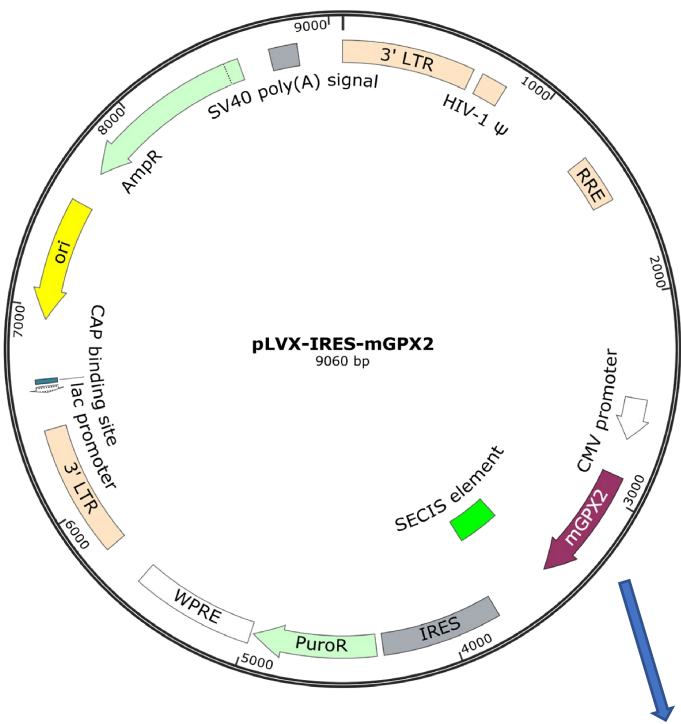
Panel	Antibody	Source	Catalogue #	Fluorophore	CD4+ T cells	CD8+ T cells	NK cells	B cells	G-MDSC	Monocytes	MΦ	M1 MΦ	M2 MΦ	Dendritic cells
Lymphoid	Anti-mouse CD45 (30-F11)	ThermoFisher	47-0451-82	APC-EF780	+	+	+	+						
Lymphoid	Anti-mouse TCRB (H57-597)	ThermoFisher	15-5961-82	APC	+	+	+							
Lymphoid	Anti-mouse CD4 (GK1.5)	ThermoFisher	15-0041-82	PE-CYN5		+	+							
Lymphoid	Anti-mouse CD8A (KT15)	Santa Cruz	SC-53473	PE			+							
Lymphoid	Anti-mouse FOXP3 (FJK-16S)	ThermoFisher	56-5698-82	FITC			+							
Lymphoid	Anti-NK1.1 (PK136)	ThermoFisher	45-5941-82	PerCP-CYN5.5				+						
Lymphoid	Anti-CD19	ThermoFisher	48-0193-82	EF450				+						
Meyloid	Anti-mouse CD45 (30-F11)	ThermoFisher	47-0451-82	APC-EF780					+	+	+	+	+	+
Meyloid	Anti-mouse CD11b (M1/70)	BD Pharmingen	561960	APC					+	+	+	+	+	
Meyloid	Anti-mouse LY-6G (RB6-8C5)	ThermoFisher	12-5931-82	PE					+					
Meyloid	Anti-mouse LY6C (HK1.4)	ThermoFisher	45-5932-82	PerCP-CYN5.5						+				
Meyloid	Anti-mouse MHCII (IA/1E)	ThermoFisher	56-5321-82	AF700					+		+	+	+/-	+
Meyloid	Anti-mouse F4/80 (BM8)	ThermoFisher	48-4801-82	EF450							+	+	+	
Meyloid	Anti-mouse CD11C (N418)	ThermoFisher	15-0114-82	PE-CYN5										+
Meyloid	Anti-iNOS/NOS	BD Pharmingen	610330	FITC								+		

The table above lists the antibodies to markers used in the immune panels along with which leukocyte

subsets they identify. Both panels included anti-CD45 (pan leukocyte) antibody and DAPI stain to exclude dead cells from analysis.

Supplementary References

1. Veeramachaneni R, Yu W, Newton JM, et al. Metformin generates profound alterations in systemic and tumor immunity with associated antitumor effects. *J Immunother Cancer* 2021;9(7) doi: 10.1136/jitc-2021-002773 [published Online First: 2021/07/08]
2. Newton JM, Hanoteau A, Sikora AG. Enrichment and Characterization of the Tumor Immune and Non-immune Microenvironments in Established Subcutaneous Murine Tumors. *J Vis Exp* 2018(136) doi: 10.3791/57685
3. Frederick M, Skinner HD, Kazi SA, et al. High expression of oxidative phosphorylation genes predicts improved survival in squamous cell carcinomas of the head and neck and lung. *Sci Rep* 2020;10(1):6380. doi: 10.1038/s41598-020-63448-z
4. Liberzon A, Birger C, Thorvaldsdóttir H, et al. The Molecular Signatures Database (MSigDB) hallmark gene set collection. *Cell Syst* 2015;1(6):417-25. doi: 10.1016/j.cels.2015.12.004
5. Monti S, Tamayo P, Mesirov J, et al. Consensus Clustering: A Resampling-Based Method for Class Discovery and Visualization of Gene Expression Microarray Data. *Machine Learning* 2003;52(1):91-118. doi: 10.1023/A:1023949509487



Murine GPX2

MAYIAKSFYDLSAVGLDGEKIDFNTFRGRAVLIENVASLU GTTTRDYNQLNELQCRFPRLVVLGFCNQFGHQENCQN
EEILNSLKYYVRPGGGYQPTFSLTQKCDVNGQNEHPVFAYLKDKLPYPYDDPFSLMTDPKLIWSPVRRSDVSWNF EKFLI
GPEGEPFRRYSRSFQTINIEPDIKRLKVAI*

U = Selenocysteine

Figure S1. Murine GPX2 expression construct. Murine cDNA encoding GPX2 and the 3' SECIS regulatory sequence for incorporation of the SEC amino acid were cloned into the pLVX-IRES-Puro construct which lacks tags besides the puromycin resistance gene. The predicted amino acid sequence containing SEC (i.e. U) appears underneath.

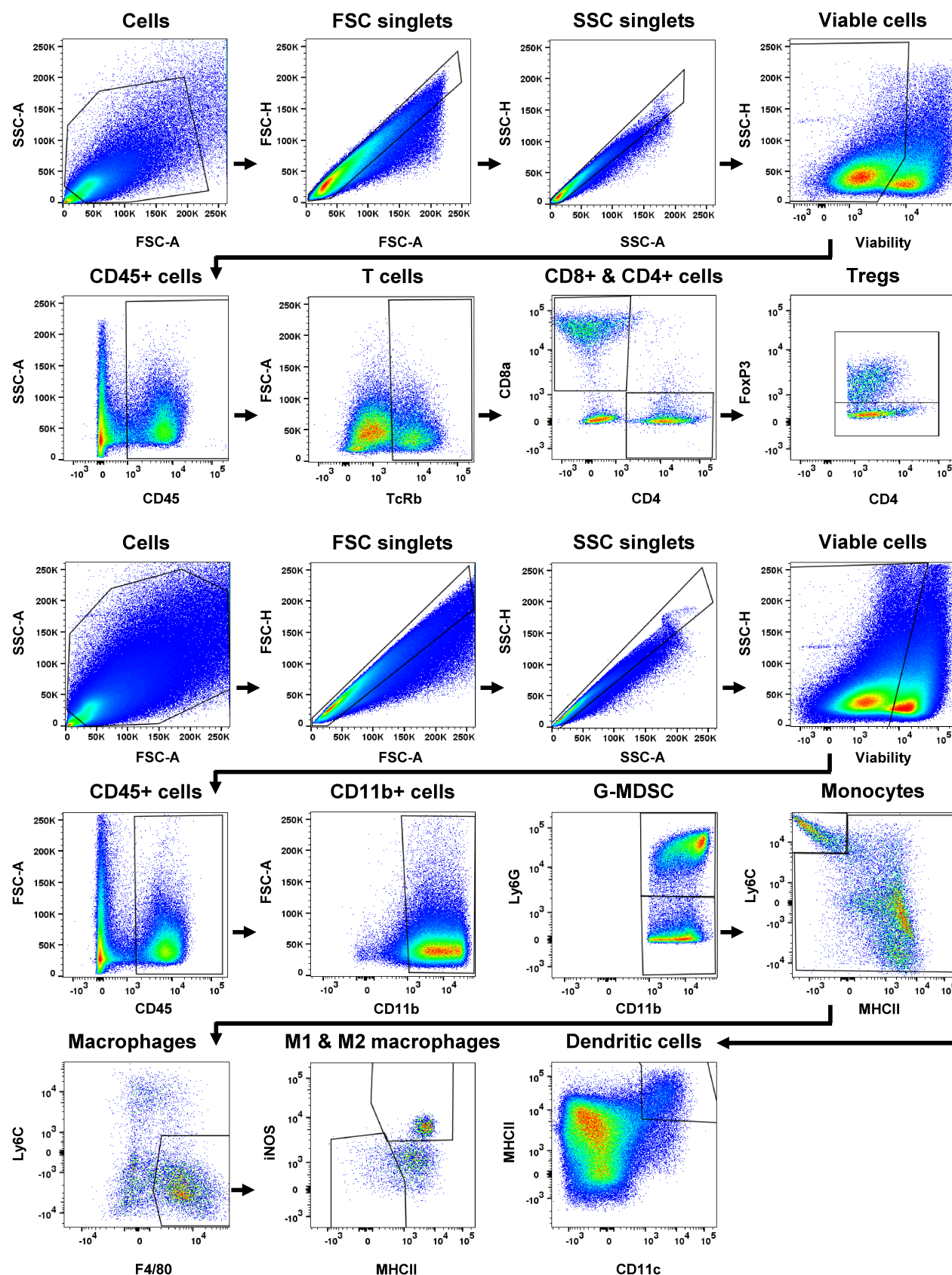


Figure S2. Gating strategy for *in vivo* tumor immune profiling by flow cytometry. Single cell suspensions prepared from mouse tumors and enriched for leukocytes as described in methods were stained with multiple fluorescently tagged antibodies recognizing molecules part of either a lymphocyte or myeloid panel. After setting appropriate forward and side scatter gates, only viable (i.e., DAPI exclusion) CD45+ leukocytes were further selected for subsequent analysis. Markers for NK and B cell subsets were included in the lymphocyte panel (Supplementary Methods) but not shown here for simplicity.

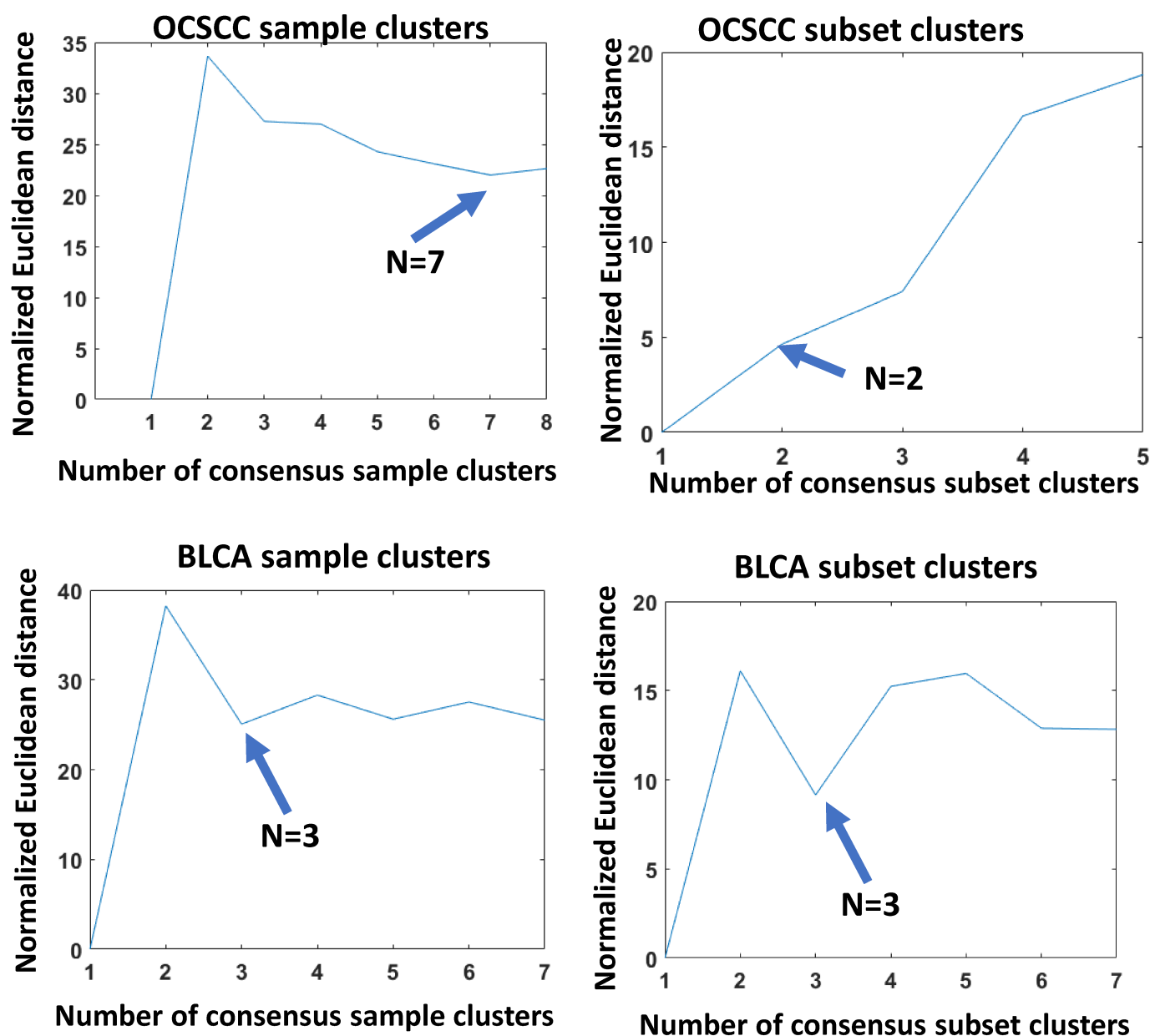
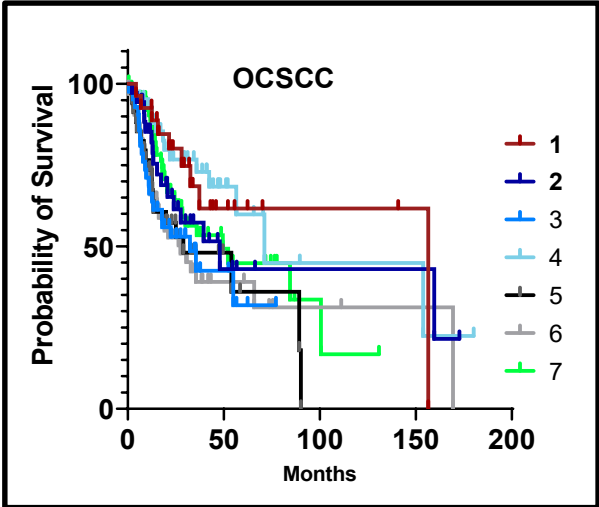
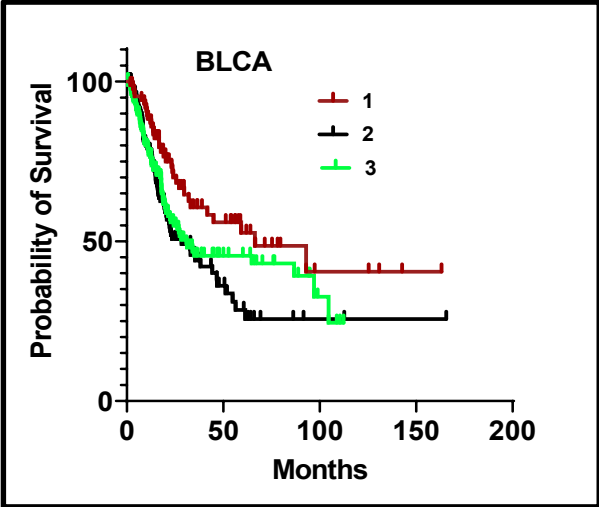


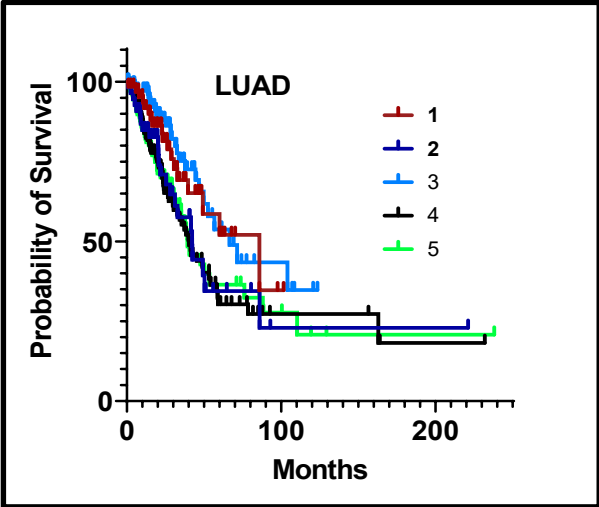
Figure S3. Strategy for choosing optimal sample and feature cluster numbers during two-way consensus hierarchical clustering. The Normalized Euclidean Distance (NED) is a measure of the difference between a transformed perfect similarity matrix and the transformed similarity matrix obtained for each value of N clusters after consensus clustering. The computationally derived similarity matrix for a given N contains the proportion “r” of times any two samples occurred in the same cluster during re-sampling for a chosen value of N clusters. Under perfect conditions this value should be equal to 1, and for two samples not found together in the same cluster the value of “r” should theoretically be 0. A simple transformation of matrix values to equal 1-r in conditions where two samples don’t co-occur in the same cluster means that the ideal transformed similarity matrix would be filled with values of 1 everywhere. The NED is taken as the square root of the residual sum of squares between a perfect transformed matrix and the actual transformed matrix, and dividing by the number of clusters (i.e. square root of the matrix size). Smaller NED values mean more robust clustering and the plots above show the tradeoff between increasing cluster complexity and gains in robustness. Example NED plots derived from clustering OCSCC and BLCA samples in one dimension or features (leukocyte subsets) in the orthogonal dimension are shown.



Cluster	1	2	3	4	5	6	7
Med Survival	156.5	48.0	32.4	71.2	29.0	27.5	49.4
P-value		0.352	0.032	0.915	0.022	0.030	0.147
Hazard ratio		1.460	2.239	0.956	2.397	2.189	1.704



Cluster	1	2	3
Med Survival	66.4	28.4	30.9
P-value		0.0042	0.043
Hazard ratio		1.817	1.515



Cluster	1	2	3	4	5
Med Survival	86.0	42.3	66.6	40.4	39.3
P-value		0.121	0.541	0.031	0.069
Hazard ratio		1.618	0.8305	1.755	1.635

Figure S4. Survival of TCGA cohort patients as a function of immunophenotype cluster. Overall survival of patients for OESophageal Squamous Cell Carcinoma (OESCC), BLCA, and LUAD stratified by their immune clusters defined previously in Figure 1. The tables indicate the median overall survival in months for each immune cluster and the P-values represent (log-rank) comparisons between survival curves using patients with the highest degree of leukocyte infiltration as the reference. LUSC is not shown because none of the comparisons were statistically significant.

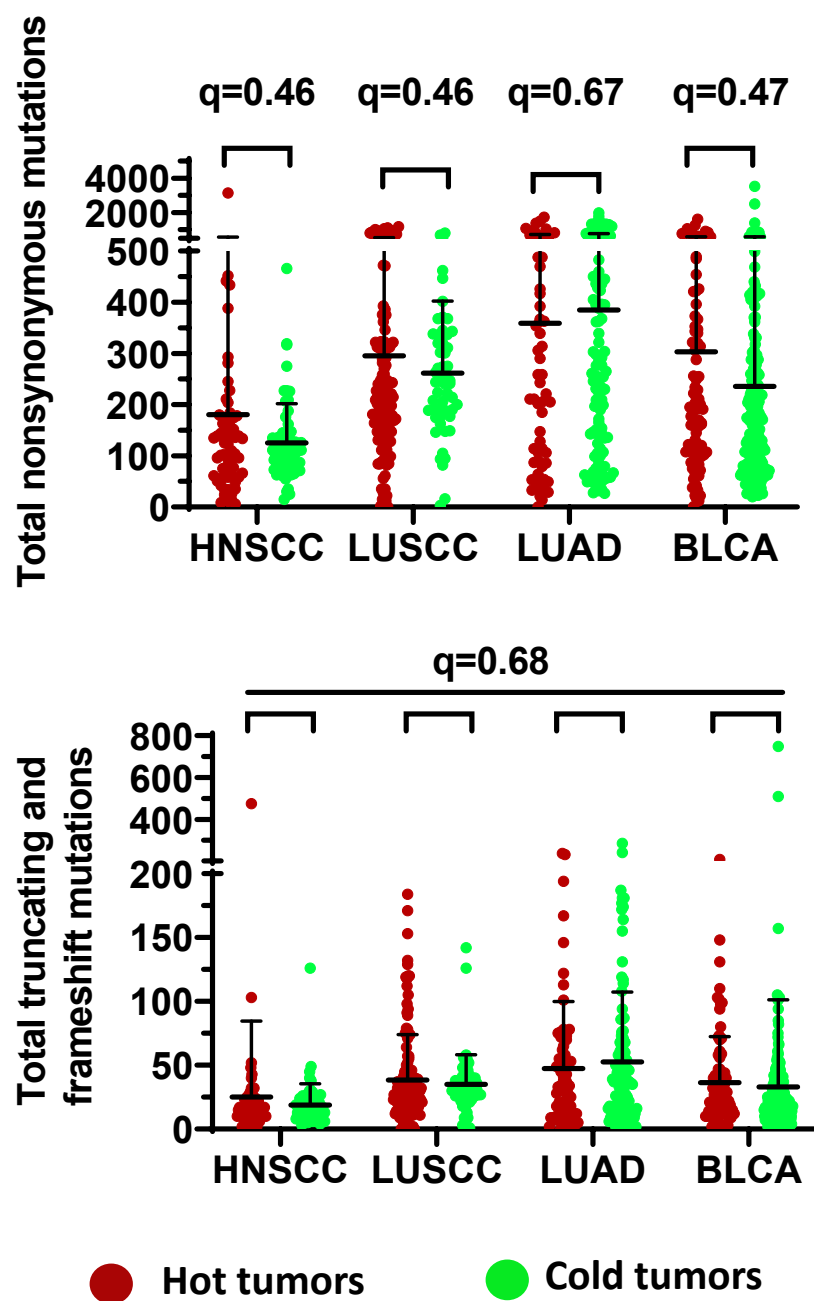


Figure S5. Immune phenotype is unrelated to tumor mutation burden. The number of total non-synonymous mutations or truncating mutations (nonsense, frameshift, splice site,) that also included stop loss mutations that add new amino acids, were compared among hot and cold tumors after correcting for multiple comparisons. No significant differences or q-values were found for comparisons in any of the tumor types.

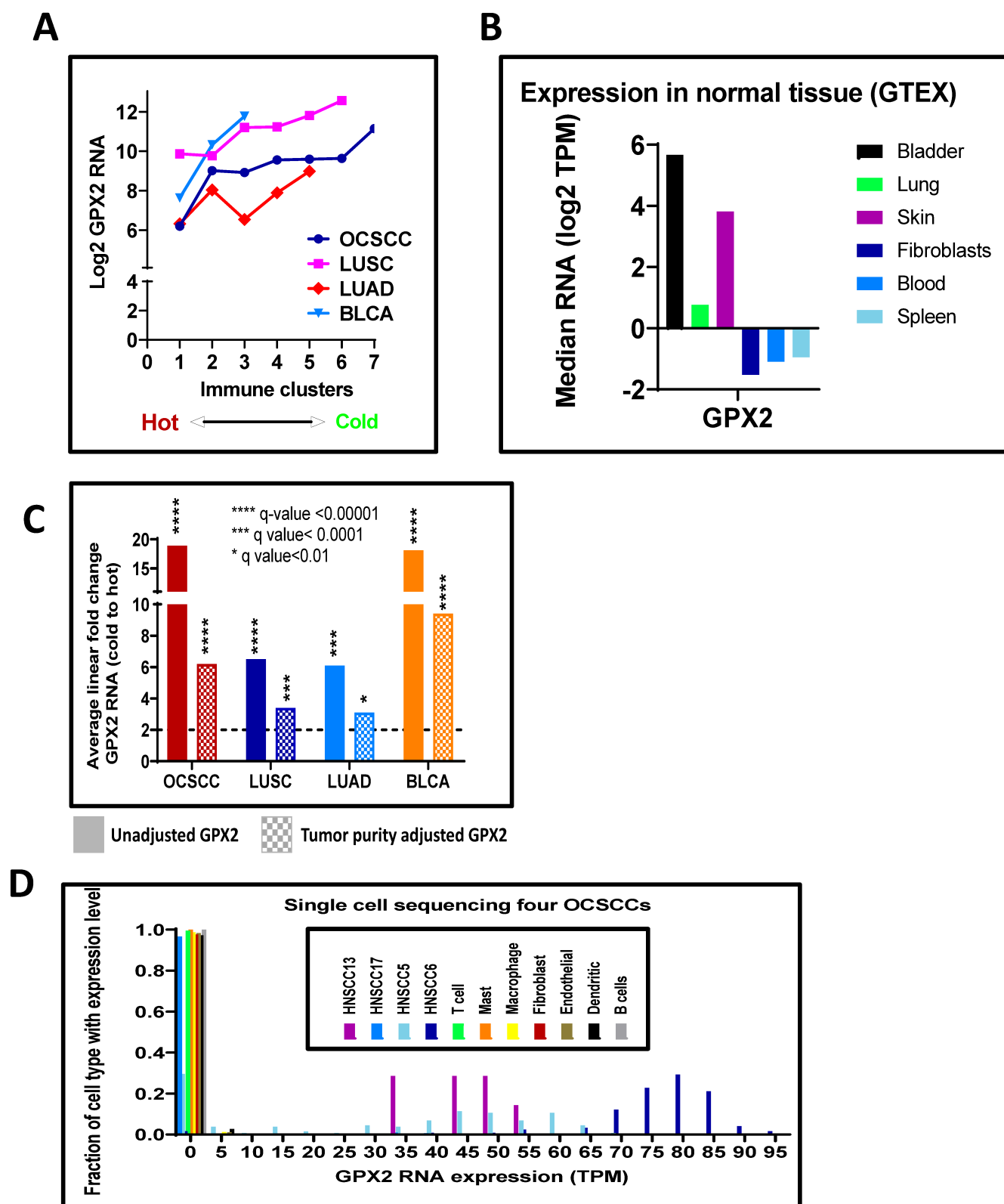


Figure S6. GPX2 expression correlates with immunophenotype across four smoking-related cancers.

(A) The average GPX2 expression value for each immune cluster (identified previously in Figure 1) is plotted for all cohorts. (B) Confirmation that GPX2 has a restricted pattern of expression in normal adult tissue from the GTEx database, characterized by relative absence in cultured fibroblasts or leukocytes present in either blood or spleen samples. (C) Statistical comparison of average GPX2 RNA levels between hot and cold tumors, with and without adjustment for tumor purity. (D) Histogram of single cell RNA-seq data from four OCSCC downloaded from GEO (GSE103322) demonstrate expression of GPX2 is predominately from tumor cells and not stromal compartments.

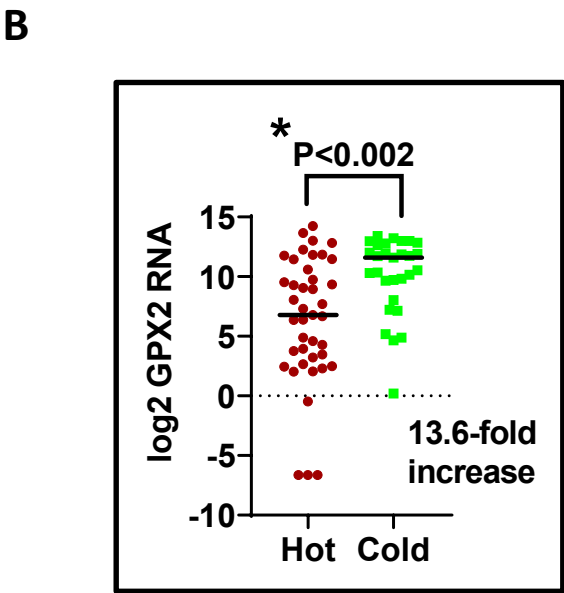
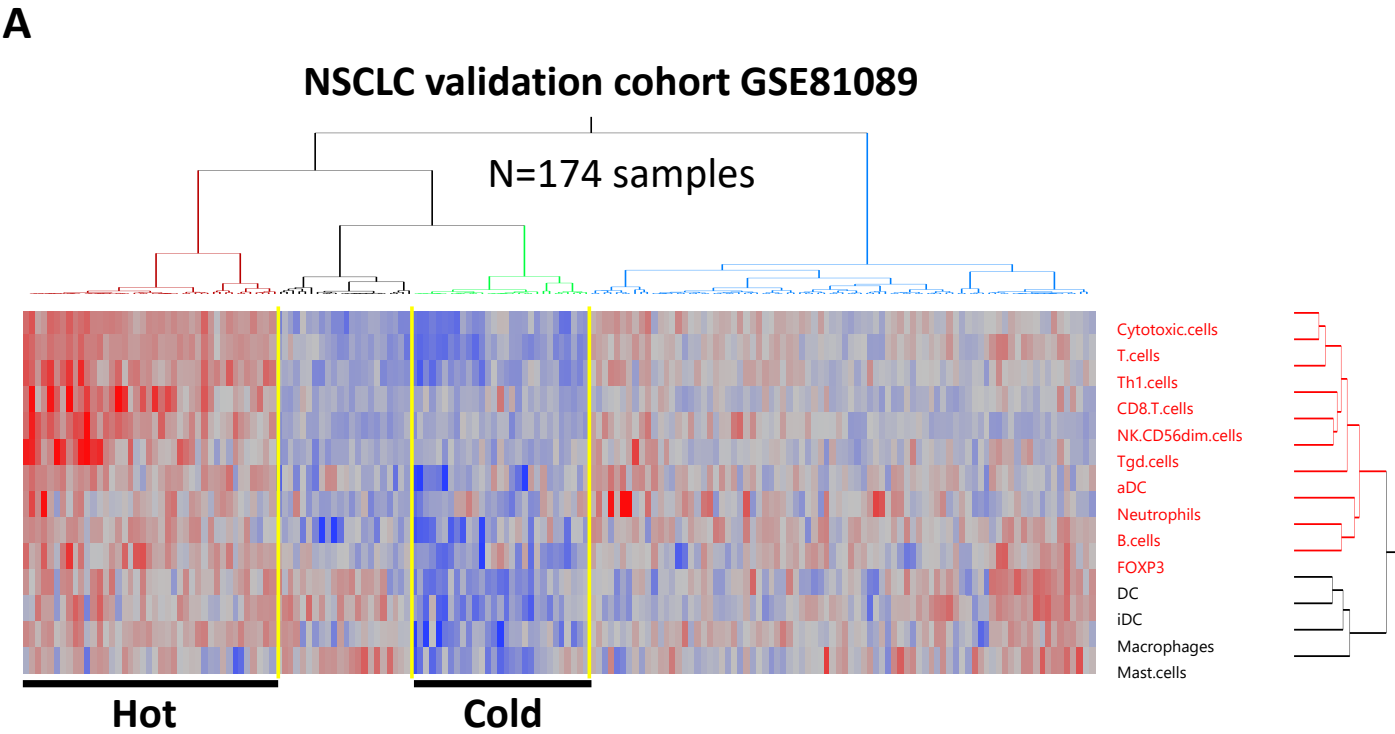


Figure S7. GPX2 overexpression is observed in cold tumors from a TCGA-independent cohort of lung tumors. RNA expression values from a Swedish cohort of lung tumors (GSE81089) consisting of 67 squamous and 107 unspecified histologies, but excluding large cell carcinomas, were converted to FPKM, upper quartile normalized, and used in subsequent ssGSEA analysis (A) and for comparison of GPX2 levels among hot and cold tumors (B).

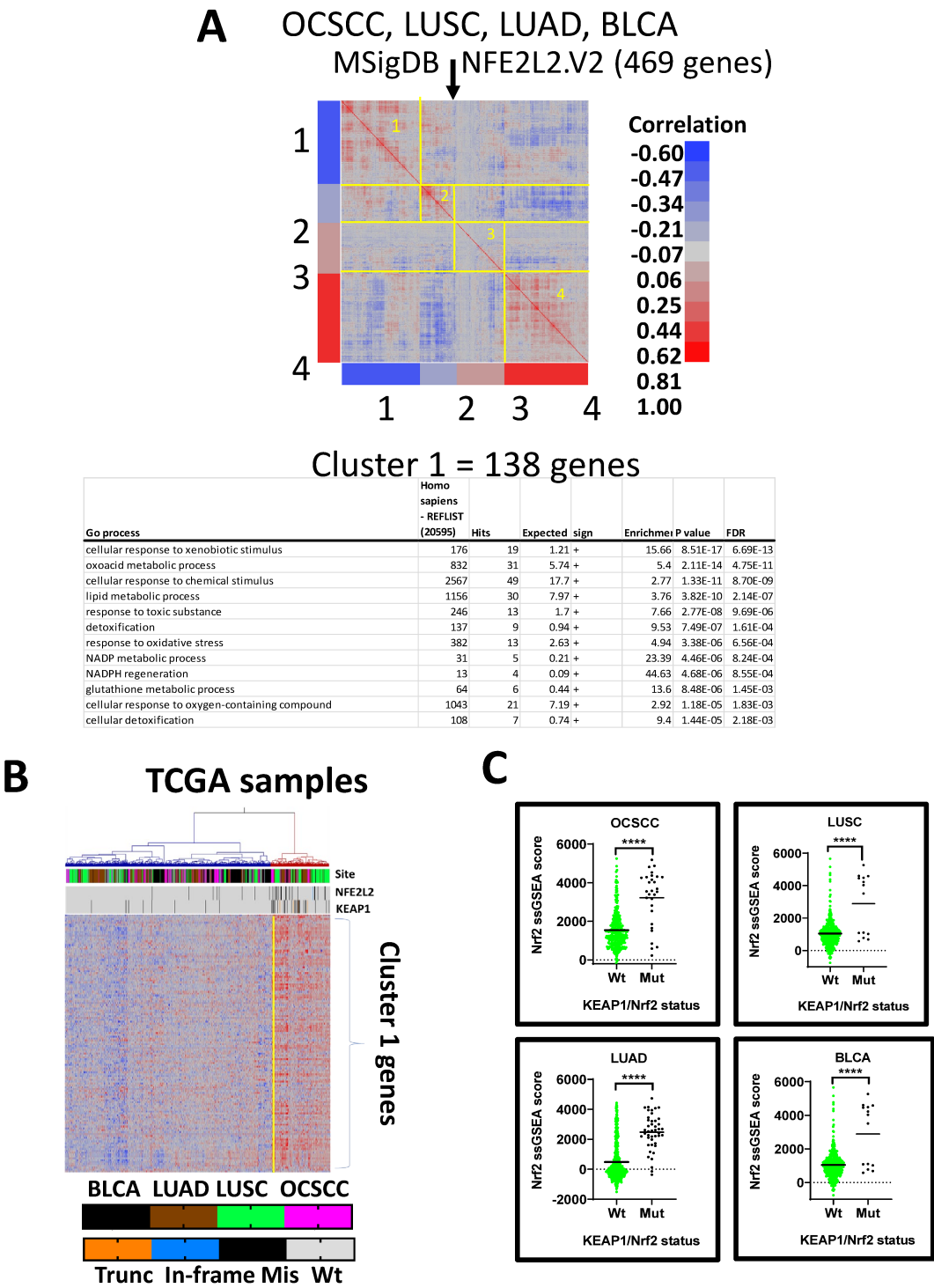


Figure S8. Validation of a gene signature for Nrf2 activation. (A) RNA from expression of 469 genes obtained from the NFE2L2.V2 gene list (i.e., candidate Nrf2 downstream targets) were examined for their cross-correlation among a pooled cohort of TCGA samples encompassing OCSCC, LUSC, BLCA, and LUAD. Cross correlation values were used for hierarchical clustering to identify the 4 gene modules along the diagonal and GO enrichment analysis was performed separately on each module to identify the one most likely representing the oxidative stress response of Nrf2. GO analysis identified gene cluster 1 with 138 genes (See Supplementary Tables S10 and S11), including GPX2 and the AKR1C family members, to be the best measure of Nrf2 activity using biological considerations. (B) Cohort-specific Z scores for the 138 gene signature were used to cluster TCGA samples and the cluster with greatest gene expression (red cluster) was highly enriched for tumors with known mutations in the KEAP1/NRF2 pathway. (C) Further validation of the derived Nrf2 signature showing statistically elevated Nrf2 ssGSEA scores among TCGA tumors harboring KEAP1 or NRF2 mutations in all four cohorts.

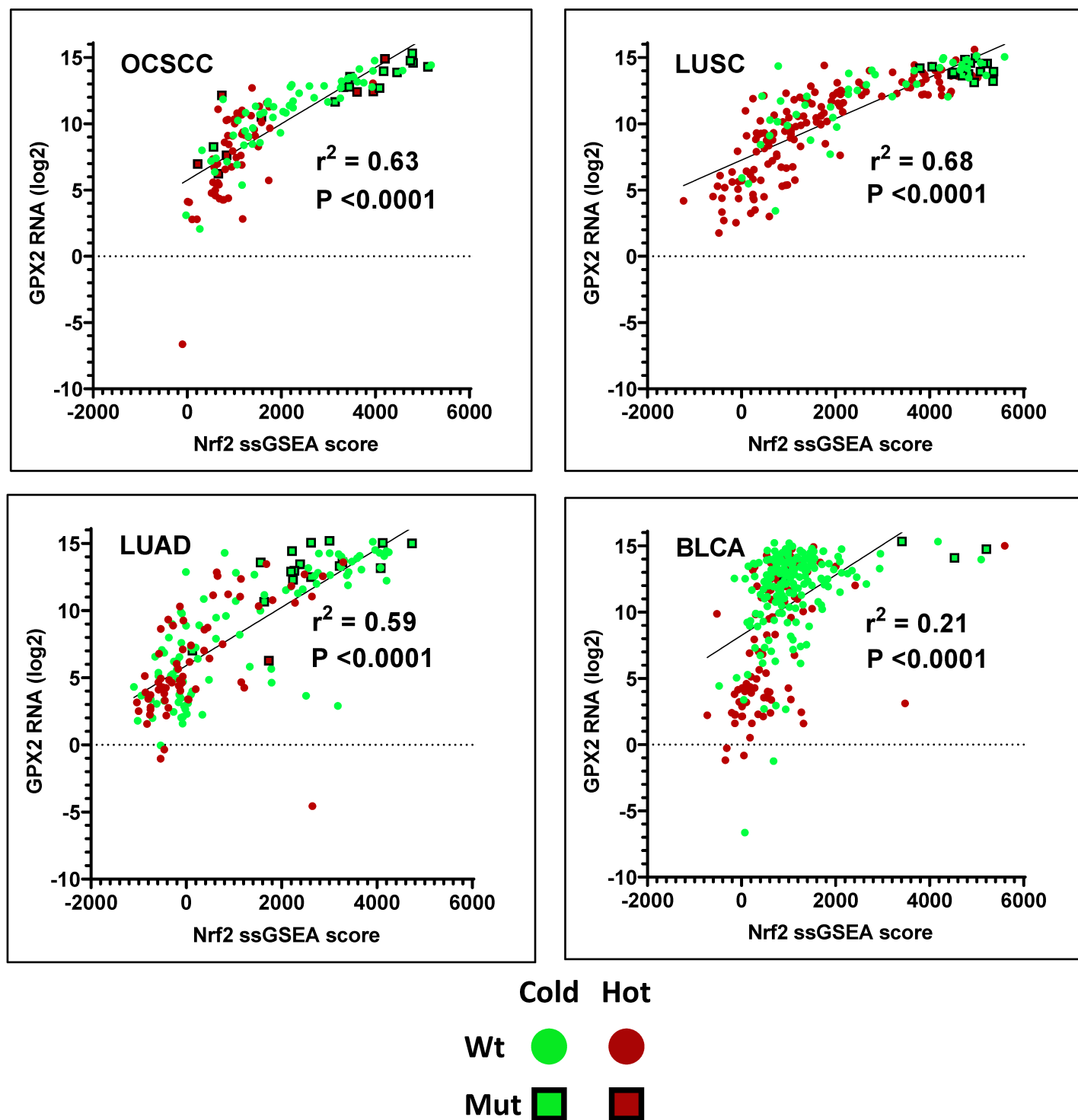


Figure S9. Expression of GPX2 RNA correlates well with Nrf2 activation in three of the smoking-related cancers. The GPX2 RNA expression was plotted versus Nrf2 ssGSEA score for individual tumors, which were visualized according to their immunophenotype (green = cold tumor; red = hot tumor) and mutational status of KEAP1/Nrf2 (circle = wild type; square = mutation). The association between GPX2 expression and Nrf2 activation status appears to hold regardless of the mutational status or immunophenotype, except in BLCA where the correlation was relatively poor.

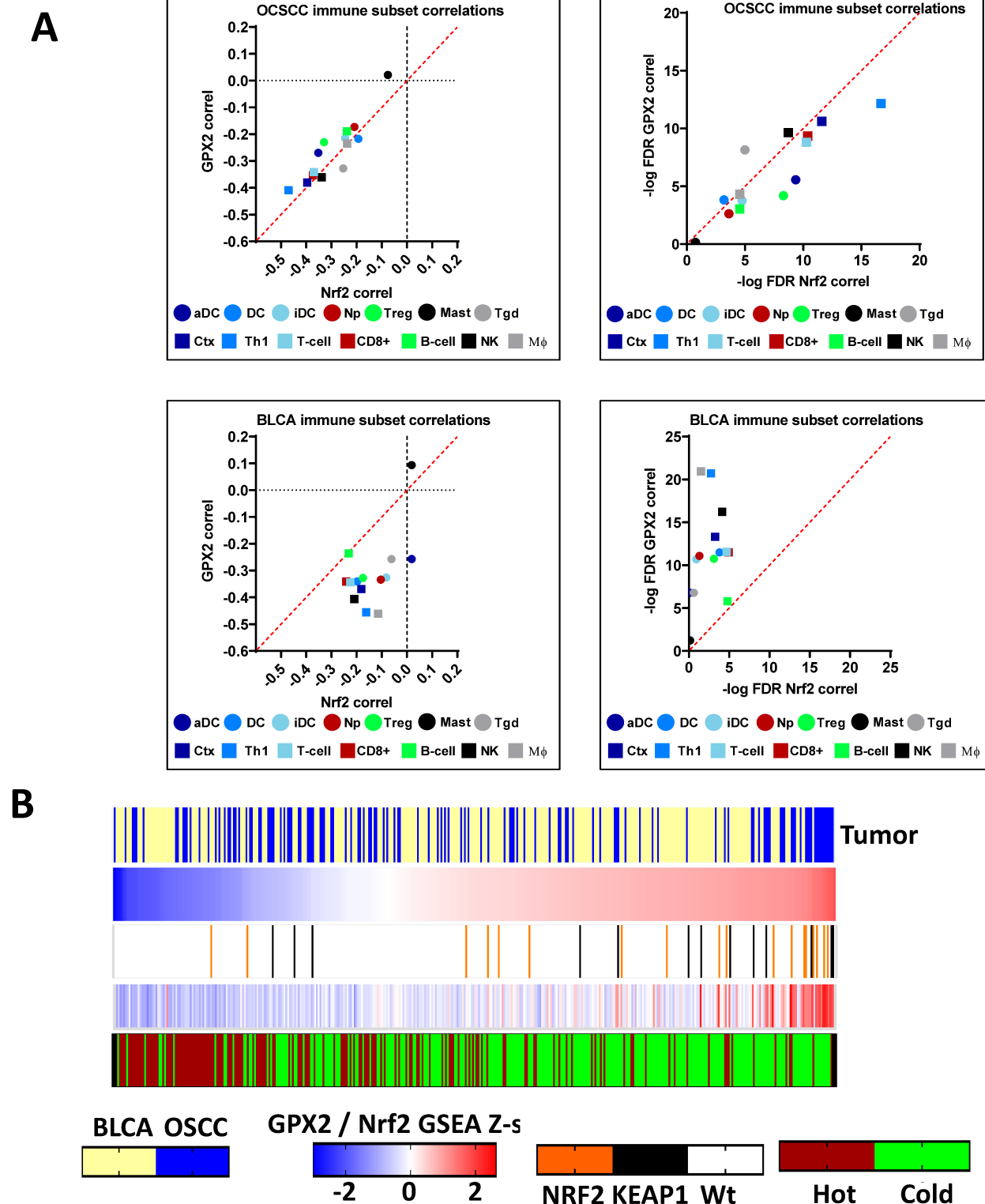


Figure S10. The TIME is more strongly associated with GPX2 than Nrf2 activation in BLCA. (A) The Pearson correlation values between individual leukocyte subtype ssGSEA scores (Supplementary Table S2) and either GPX2 RNA or Nrf2 ssGSEA are plotted for comparison (left panels) or their corresponding FDR values as a measure of significance (right panels). Deviation from the identity line (diagonal) indicates skewing or stronger association with GPX2 in the case of BLCA (lower panels). (B) Graphical representation of the relationship between immune phenotype and GPX2 expression or Nrf2 activation in the context of tumor type or KEAP1/NRF2 mutational status.

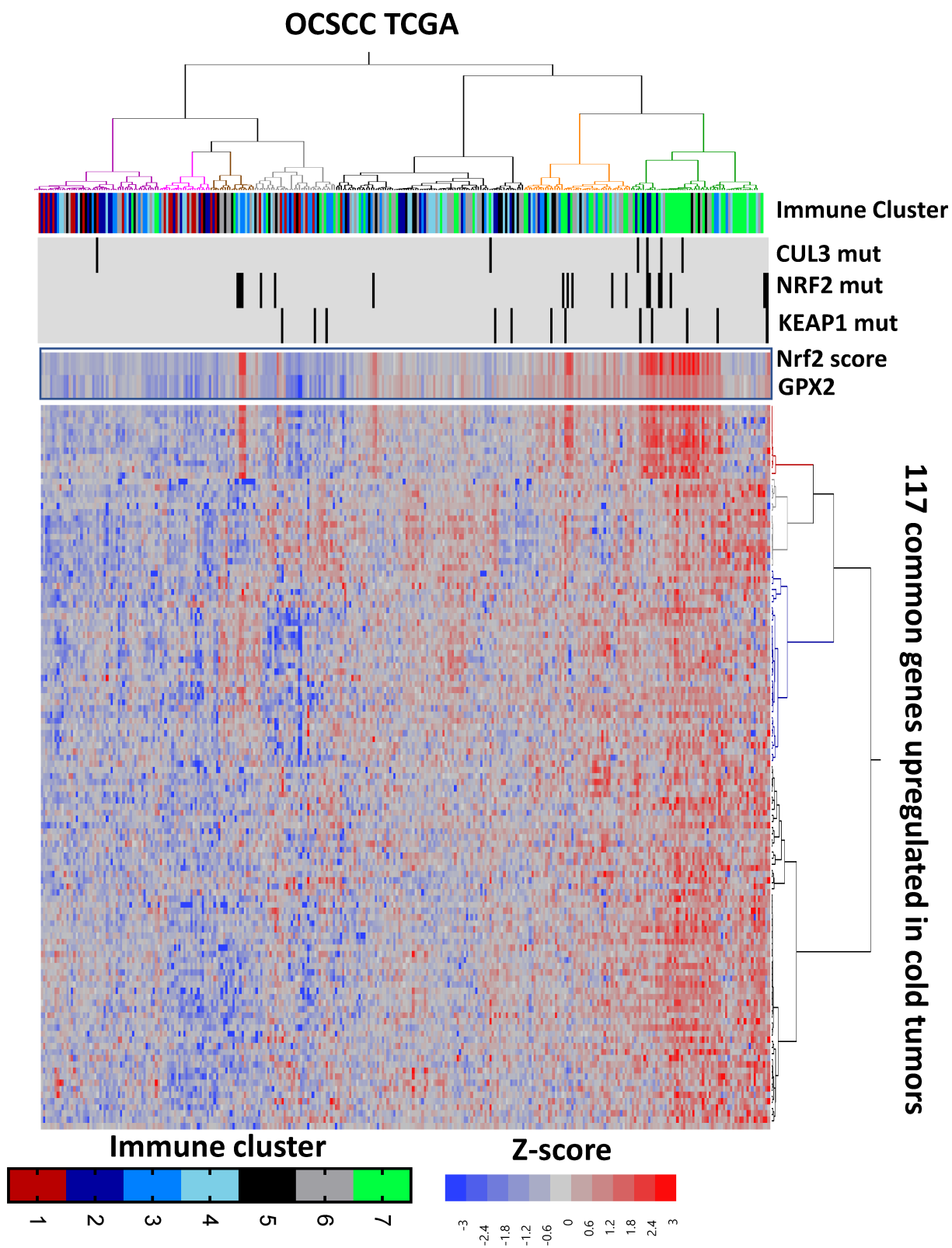
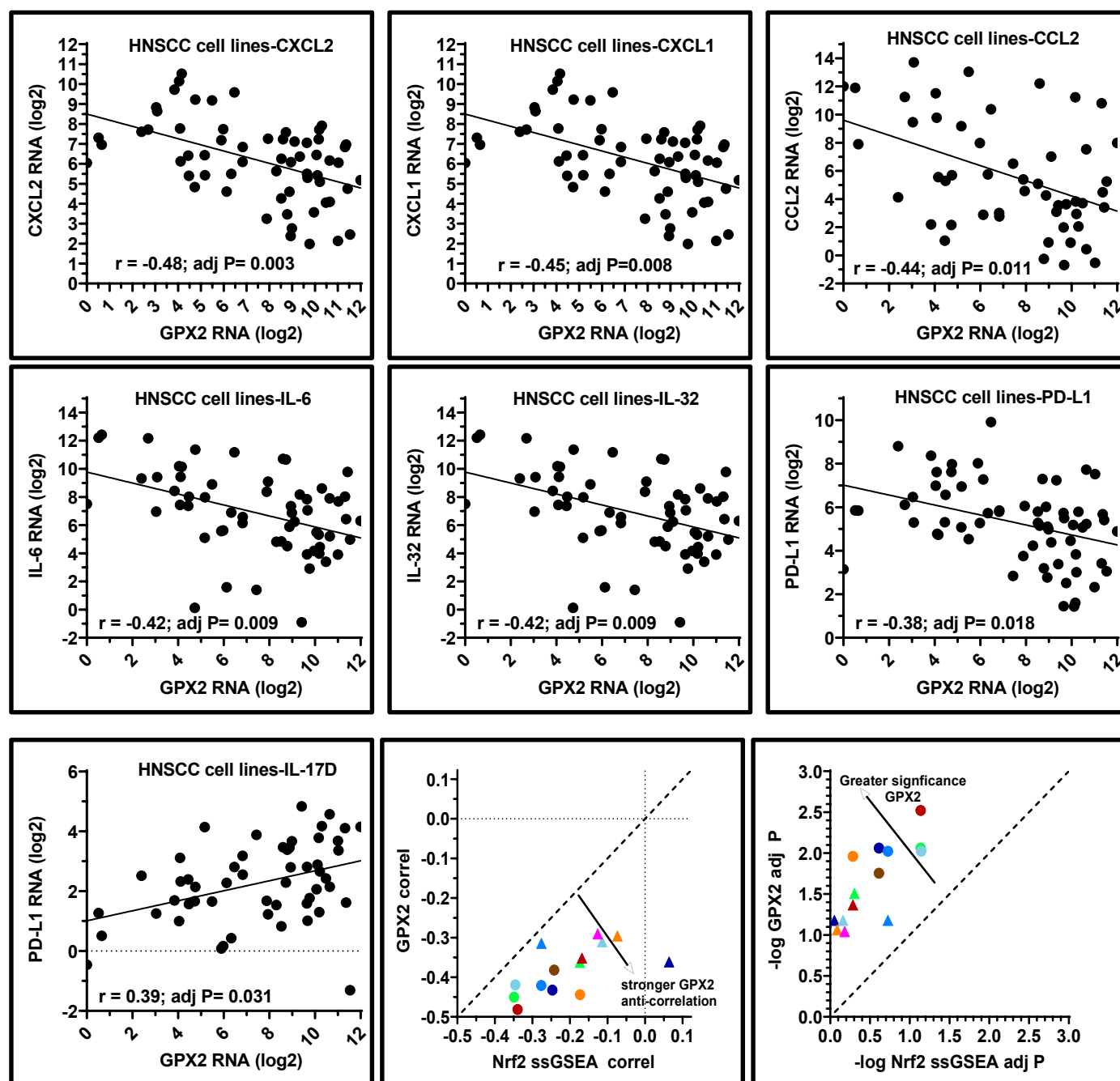
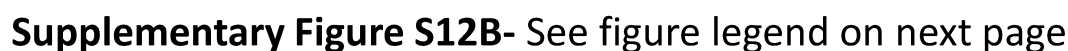


Figure S11. Evidence for Nrf2-dependent and Nrf2-independent regulation of GPX2 and other genes elevated in cold tumors. Expression of the 117 common genes upregulated in cold tumors (including GPX2) was used to cluster TCGA OCSCC samples, which were annotated by immune cluster (i.e. Figure 1), Nrf2 pathway mutational status, Nrf2 ssGSEA score or GPX2 expression. While the two clusters on the right side (orange and dark green) had the highest expression levels of the 117 common genes and contained many of the tumors with the highest levels of Nrf2 activation and GPX2 expression, a subset of cold tumors from the right-most dark green cluster were dominated by GPX2 elevation rather than Nrf2 activation suggesting a Nrf2-independent mechanism.

A



Supplementary Figure S12A- See figure legend on next page



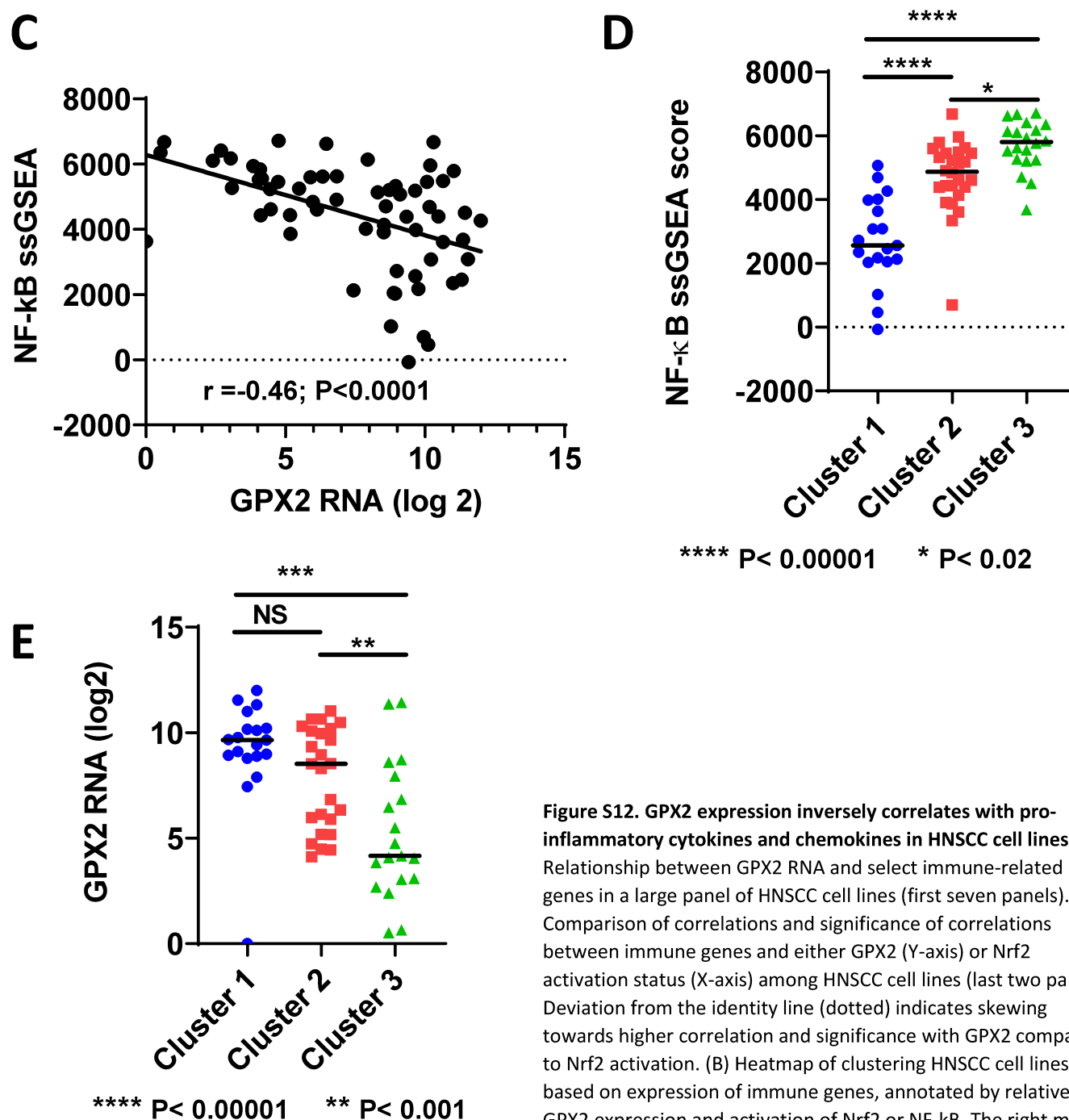


Figure S12. GPX2 expression inversely correlates with pro-inflammatory cytokines and chemokines in HNSCC cell lines. (A) Relationship between GPX2 RNA and select immune-related genes in a large panel of HNSCC cell lines (first seven panels). Comparison of correlations and significance of correlations between immune genes and either GPX2 (Y-axis) or Nrf2 activation status (X-axis) among HNSCC cell lines (last two panels). Deviation from the identity line (dotted) indicates skewing towards higher correlation and significance with GPX2 compared to Nrf2 activation. (B) Heatmap of clustering HNSCC cell lines based on expression of immune genes, annotated by relative GPX2 expression and activation of Nrf2 or NF- κ B. The right most cluster of cells (i.e., light green dendrogram) had highest levels of immune-related gene expression which was associated with higher NF- κ B activation, lower GPX2 expression, and lower Nrf2 activation. (C) Significant inverse correlation between GPX2 expression and NF- κ B activation in cell lines. (D). NF- κ B activation is significantly higher in the cell line clusters with greater cytokine/chemokine expression. (E) Conversely, GPX2 expression is significantly lower in cell line clusters with higher expression of cytokines/chemokines

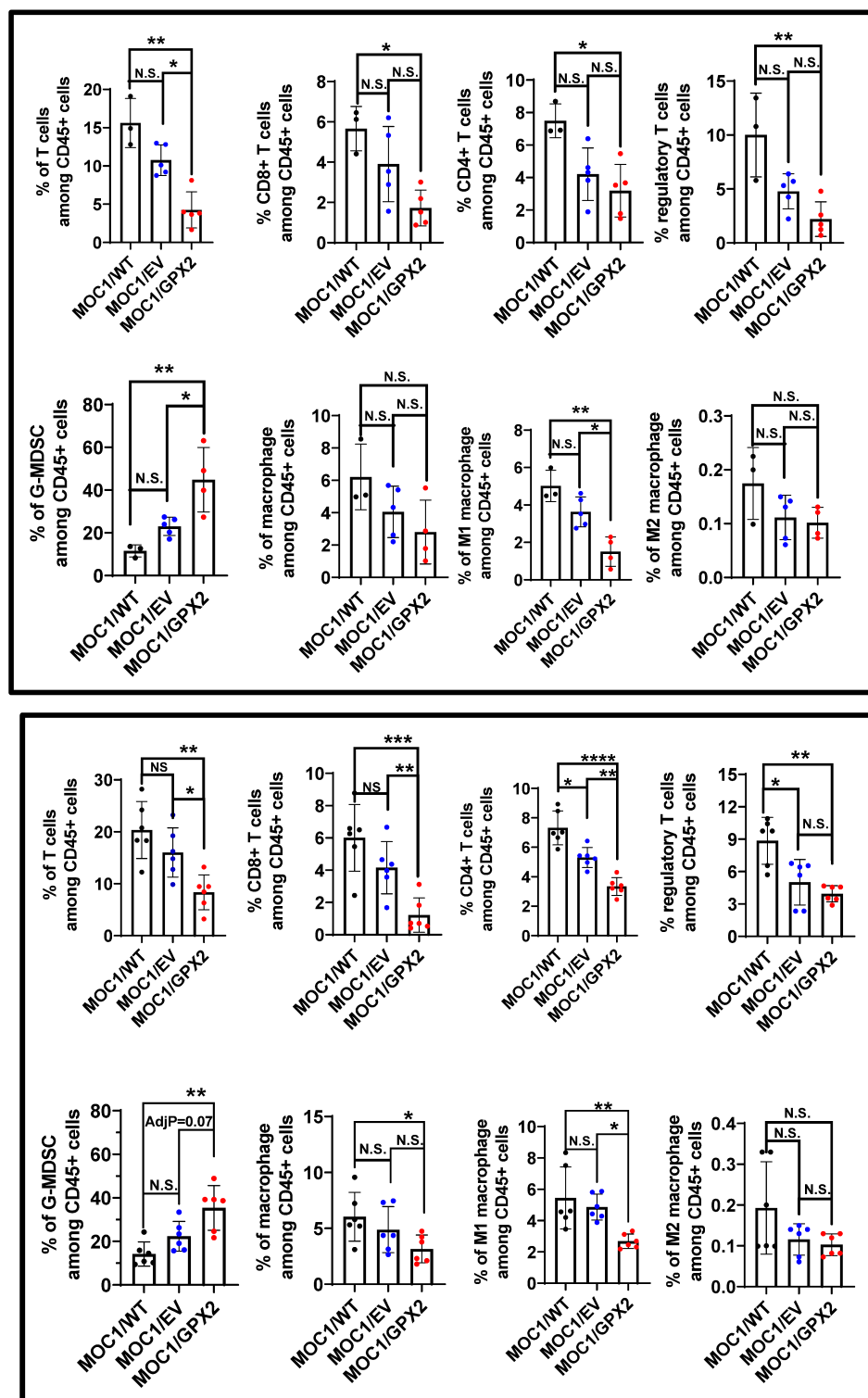
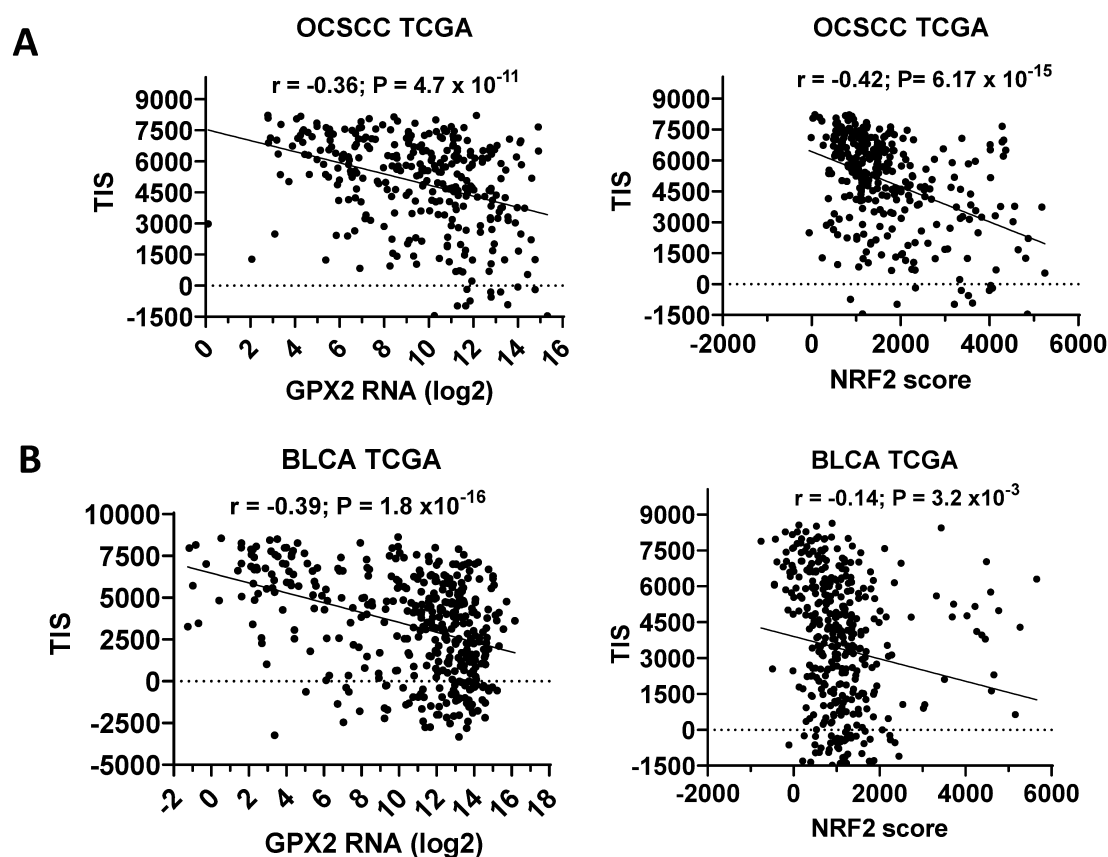


Figure S13. GPX2 drives TIME shifts. Replicate experiments (top and bottom) in which MOC1 tumors (N=3 to 5) generated from parental (WT), empty vector (EV) or GPX2 containing constructs were harvested at the same size and analyzed via flow cytometry. MOC1 parental tumors grew more slowly and had fewer tumors reach appropriate size at the time of profiling, which was done simultaneously. All data are presented as means, with error bars denoting standard deviation and percentages showed using individual circles. * denotes $p < 0.05$; ** denotes $p < 0.01$; *** denotes $p < 0.001$; **** denotes $p < 0.0001$, after performing a Tukey multiple comparison test and further adjusting P values with a Benjamini-Hochberg correction ($FDR < 0.1$) to control the family wise error within a cell line group due to multiple testing across different leukocyte subpopulations. The arcsine transformation was applied to percentages prior to statistical analysis.



TIS = Tumor inflammatory score

Figure S14. TIS correlates strongly with GPX2 RNA expression. ssGSEA using the 18 gene pan-cancer immune signature published previously by Ayers et al. [37](#) was used to correlate OCSCC (A) or BLCA (B) TIS score with expression of GPX2 RNA (left panels) or Nrf2 activation (right panels). In BLCA, the correlation with TIS score was considerably more significant with GPX2 expression compared to Nrf2 pathway status.

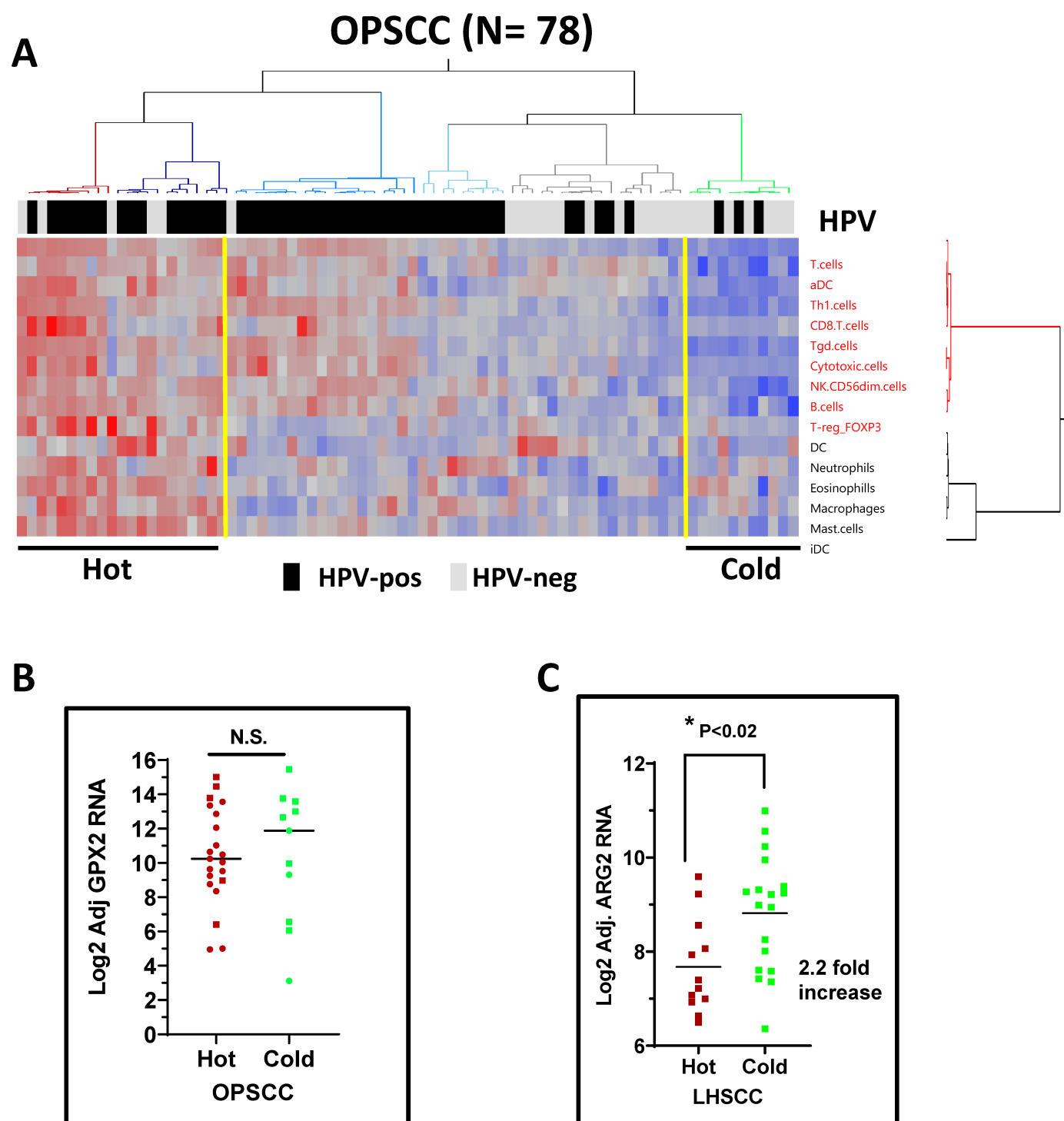


Figure S15. Analysis of cold tumors in OPSCC and LHSCC. (A) Heatmap depicting hot and cold tumors from the TCGA OPSCC samples showing that HPV, frequently present, is strongly correlated with immune infiltration ($P = 0.02$, Fisher's exact test). (B) GPX2 was not significantly correlated with immune phenotype in OPSCC, regardless of HPV status (circles = HPV-pos; squares = HPV-neg). (C) Levels of ARG2 were more than 2-fold elevated in cold LHSCC tumors after adjusting for tumor purity.

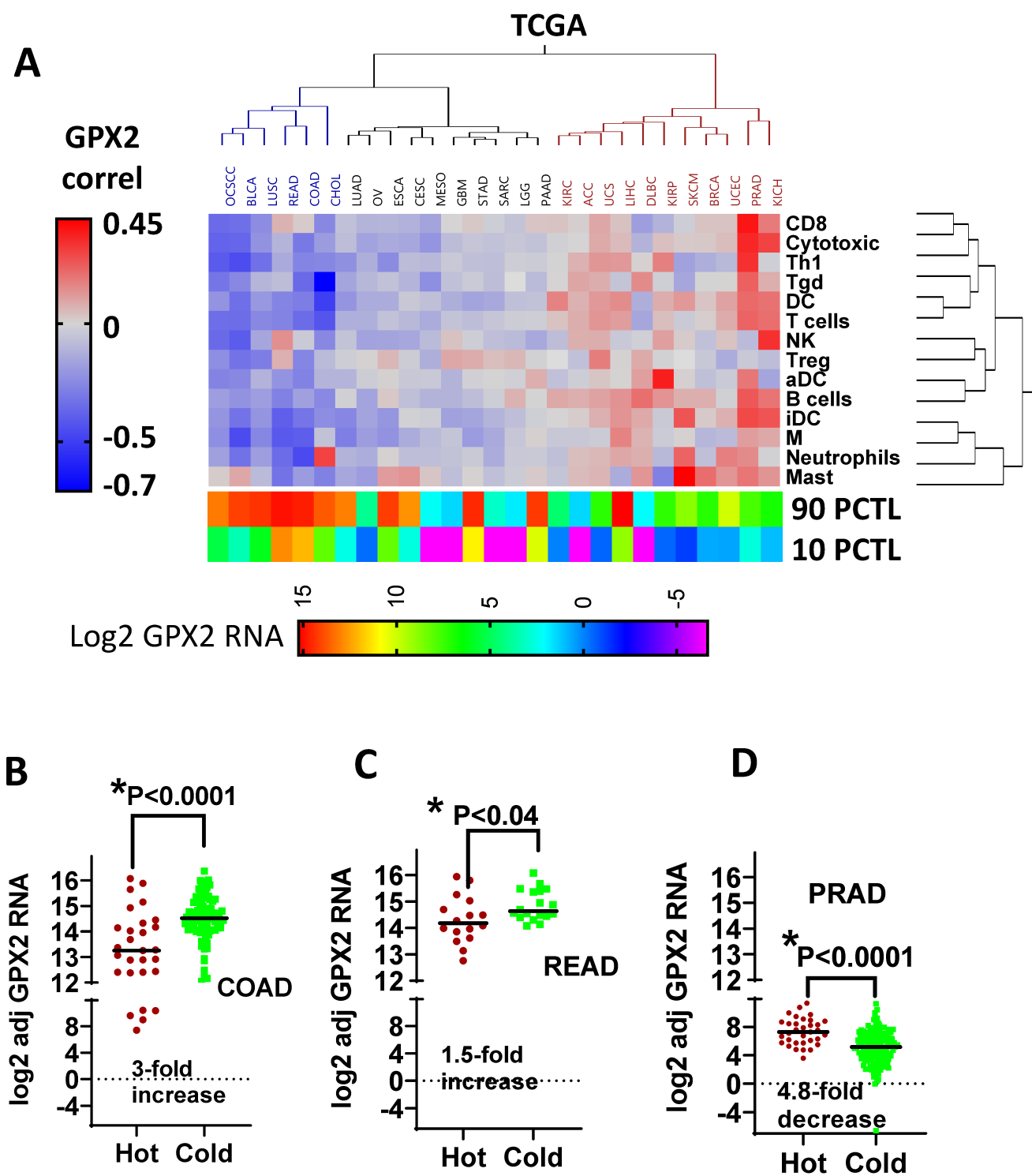


Figure S16. Pan-cancer analysis of GPX2 expression and the TIME. (A) Correlation between GPX2 RNA and degree of infiltration by leukocyte subsets (i.e., ssGSEA scores or log2 FOXP3 or Tregs) across smoking related cancers and 23 additional tumor types studied by the TCGA. Correlation values were used for 2 ways Ward’s clustering. The relative range of GPX2 expression for each cohort is represented by the 10th and 90th percentile values annotated across the bottom. Comparison of tumor purity adjusted GPX2 RNA values between hot and cold tumors for (B) colon adenocarcinoma (COAD), (C) Rectal adenocarcinoma (READ), and (D) prostate adenocarcinoma