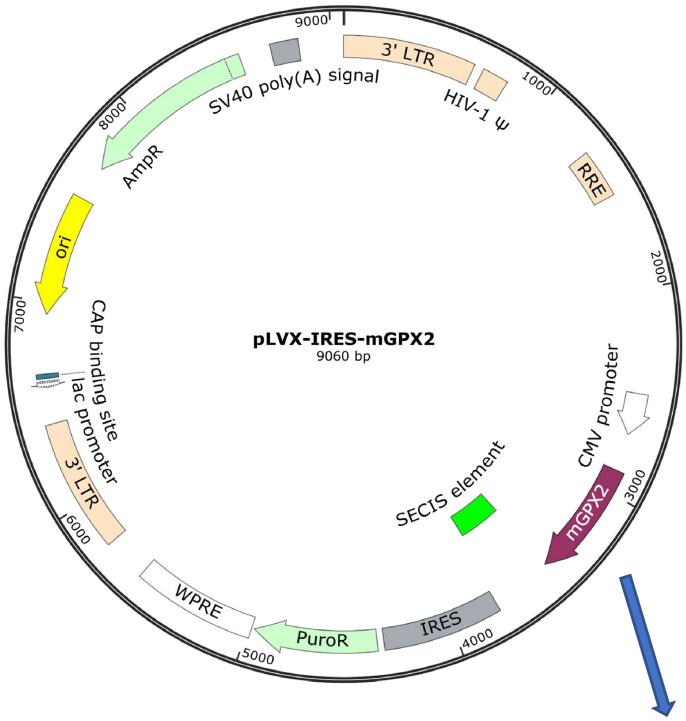


Murine GPX2

MAYIAKSFYDLSAVGLDGEKIDFNTFRGRAVLIENVASLUGTTTRDYNQLNELQCRFPRRLVVLGFPCNQFGHQENCQN
EEILNSLKYVRPGGGYQPTFSLTQKCDVNGQNEHPVFAYLKDKLPYPYDDPFSLMTDPKLIIWSPVRRSDVSWNFEKFLI
GPEGEPFRRYSRSFQTINIEPDIKRLLKVAI*

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.

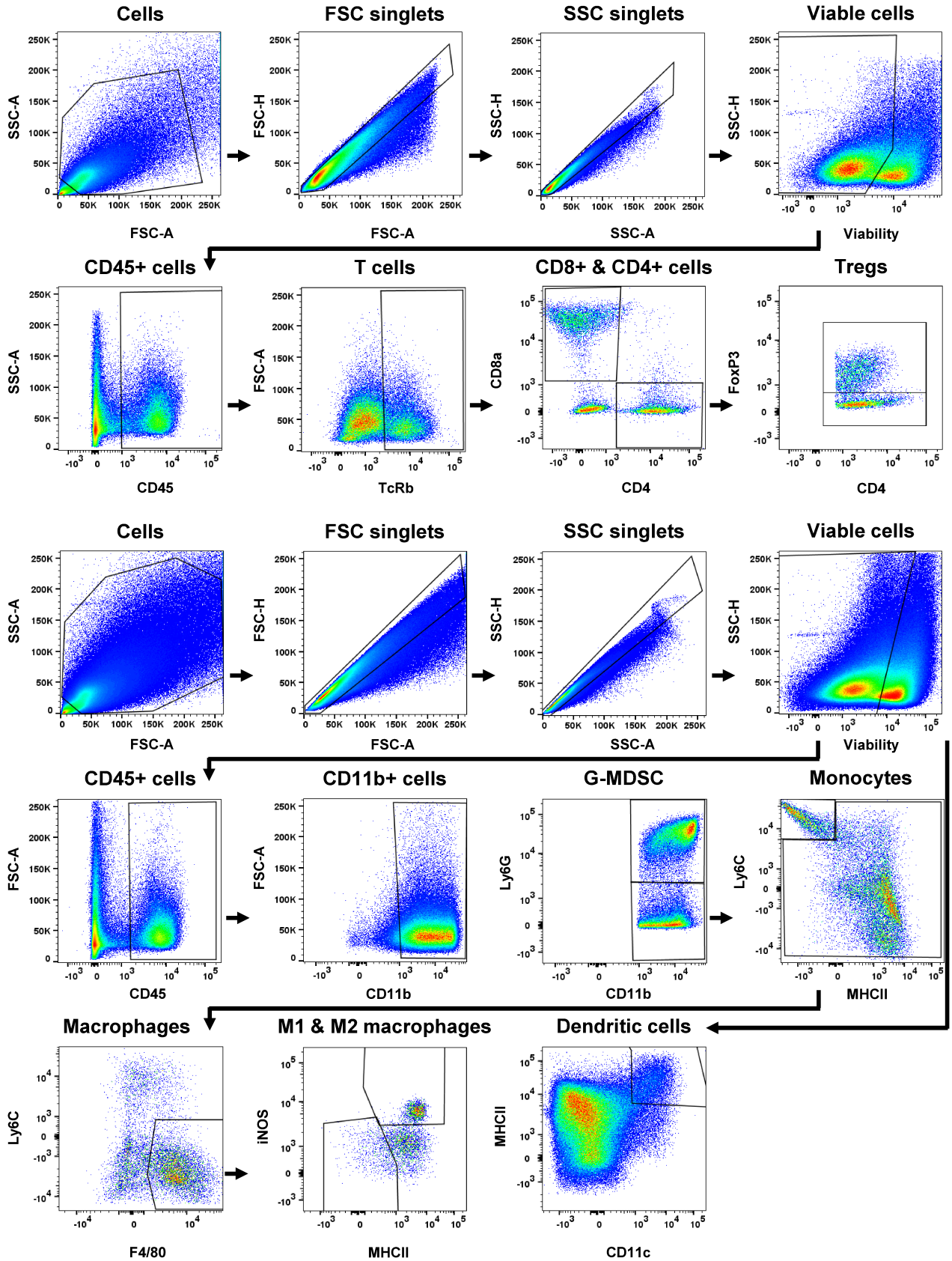


Figure S2. Gating strategy for *in vivo* tumor immune profiling by flow cytometry. Single cell suspensions prepared from mouse tumors and enriched for leukocytes as described in methods were stained with multiple fluorescently tagged antibodies recognizing molecules part of either a lymphocyte or myeloid 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.

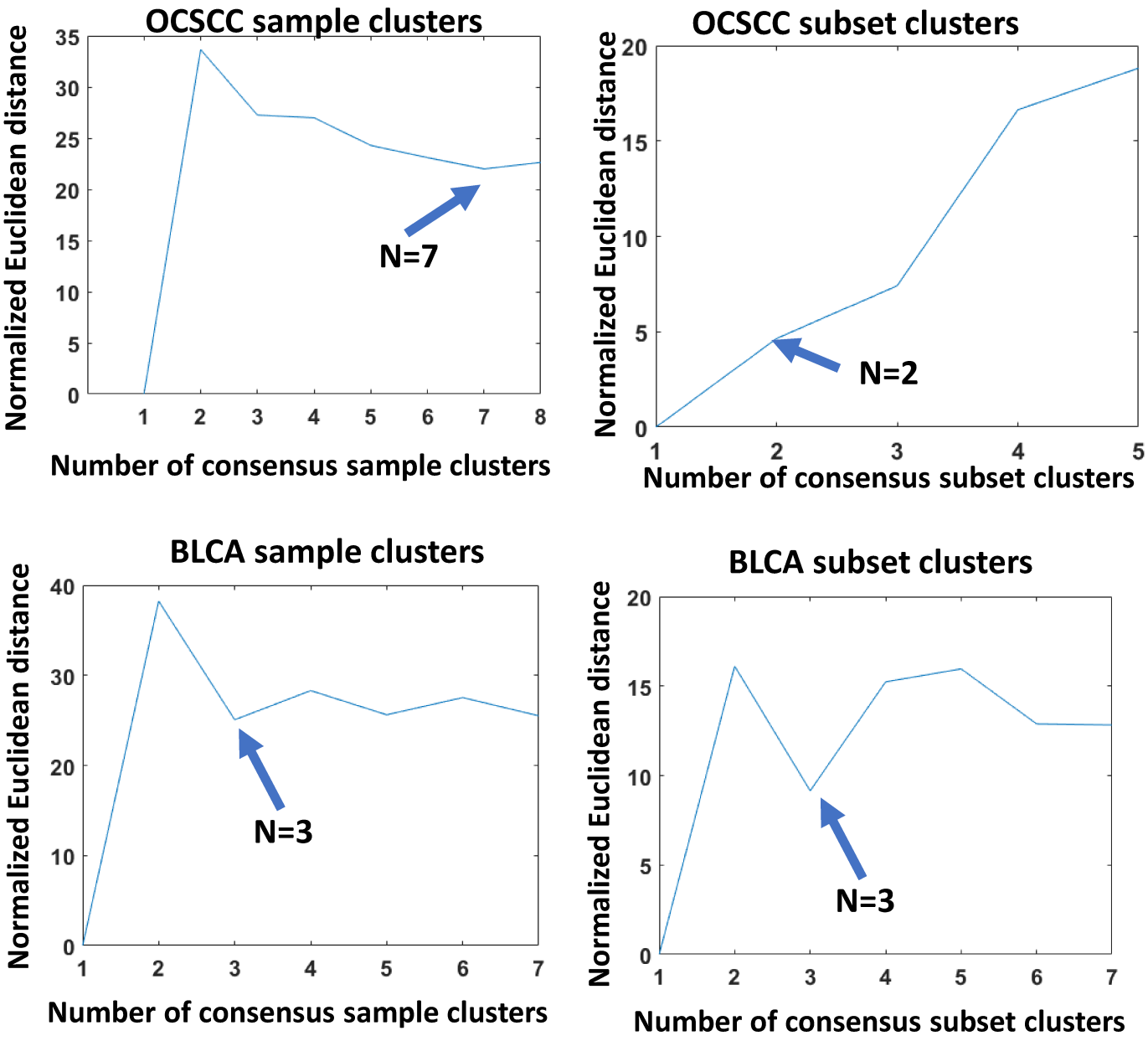
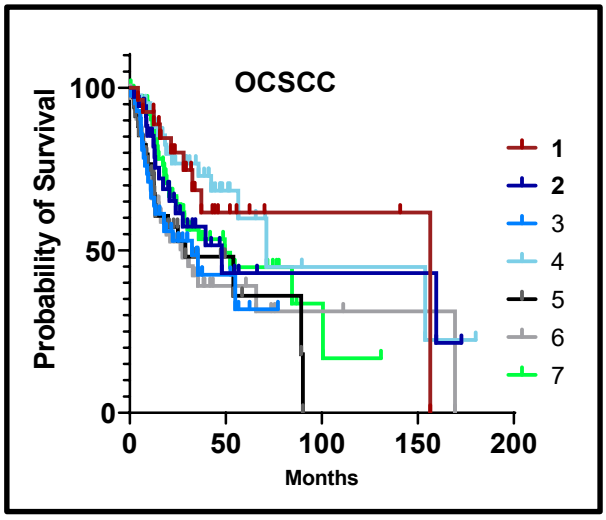
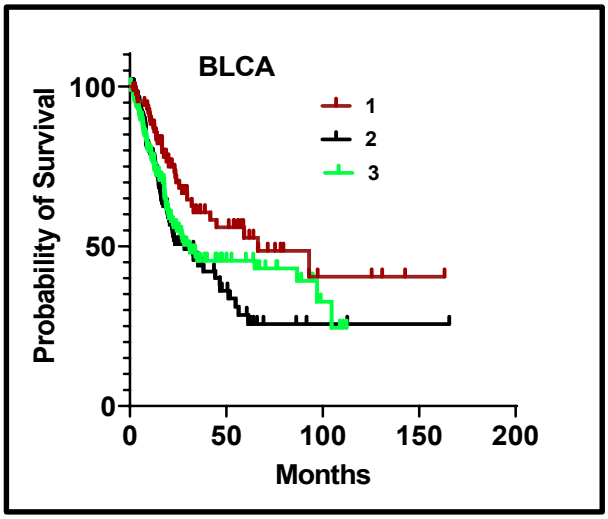


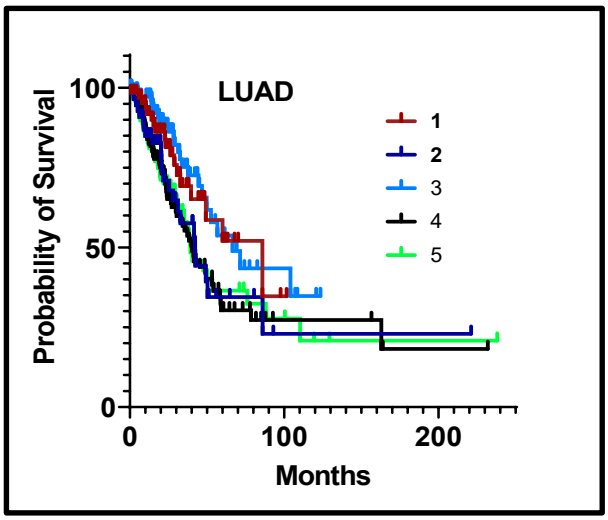
Figure S3. Strategy for choosing optimal sample and feature cluster numbers during two-way consensus hierarchical clustering. The Normalized Euclidean Distance (NED) is a measure of the difference between a transformed perfect similarity matrix and the transformed similarity matrix obtained for each 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 OCSCC, 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.

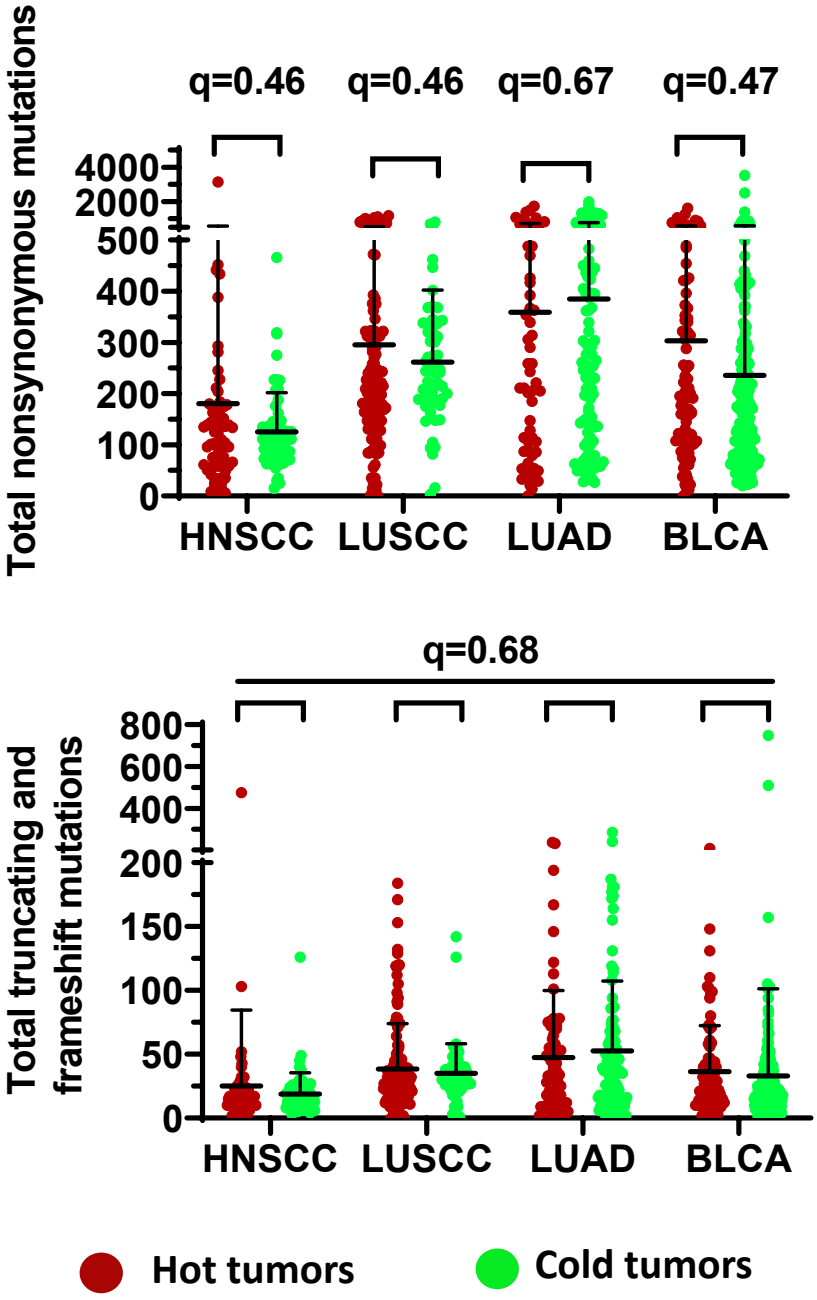


Figure S5. Immune phenotype is unrelated to tumor mutation burden. The number of total non-synonymous mutations or truncating mutations (nonsense, frameshift, splice site,) that also included stop loss mutations that add new amino acids, were compared among hot and cold tumors after correcting for multiple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

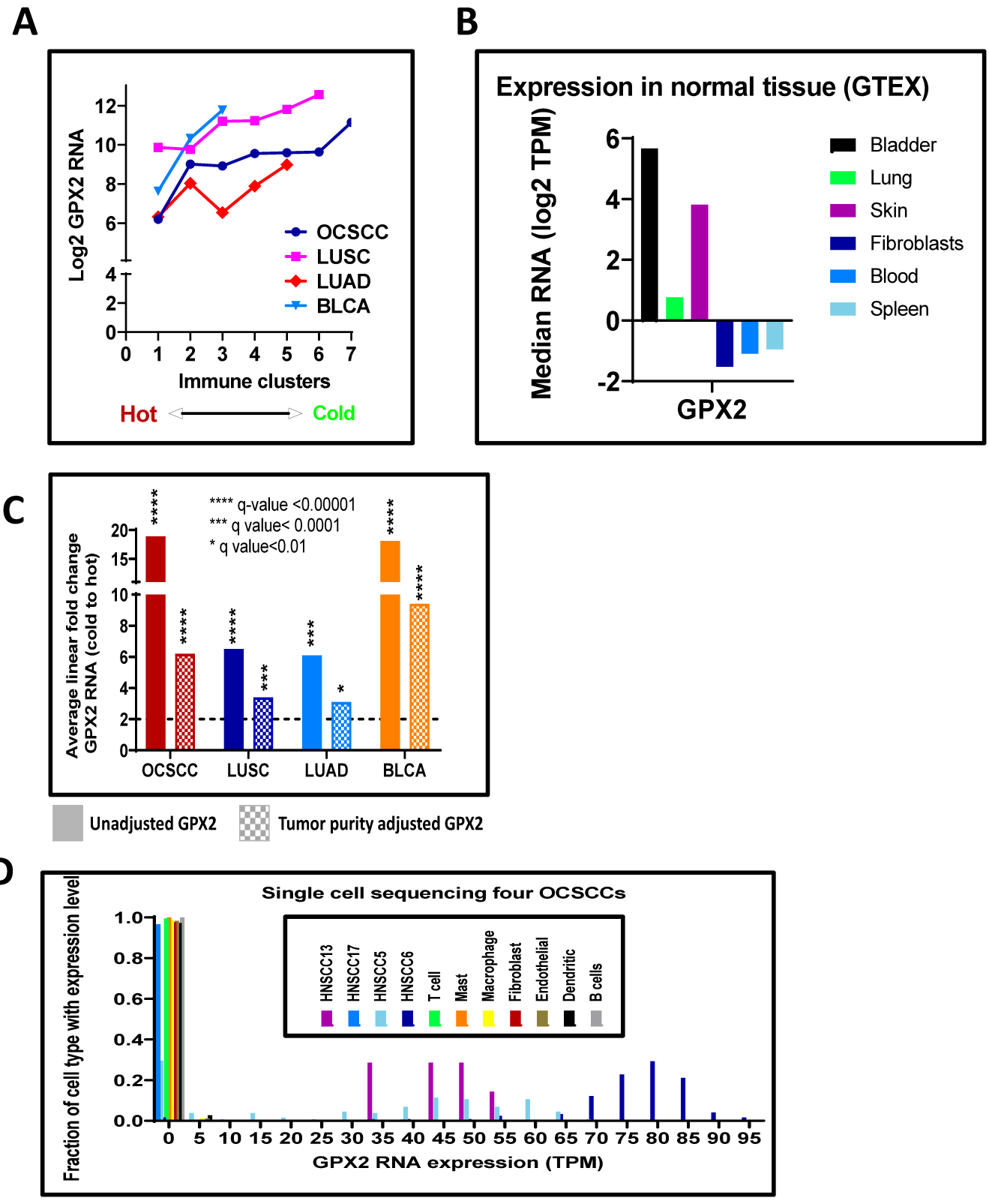
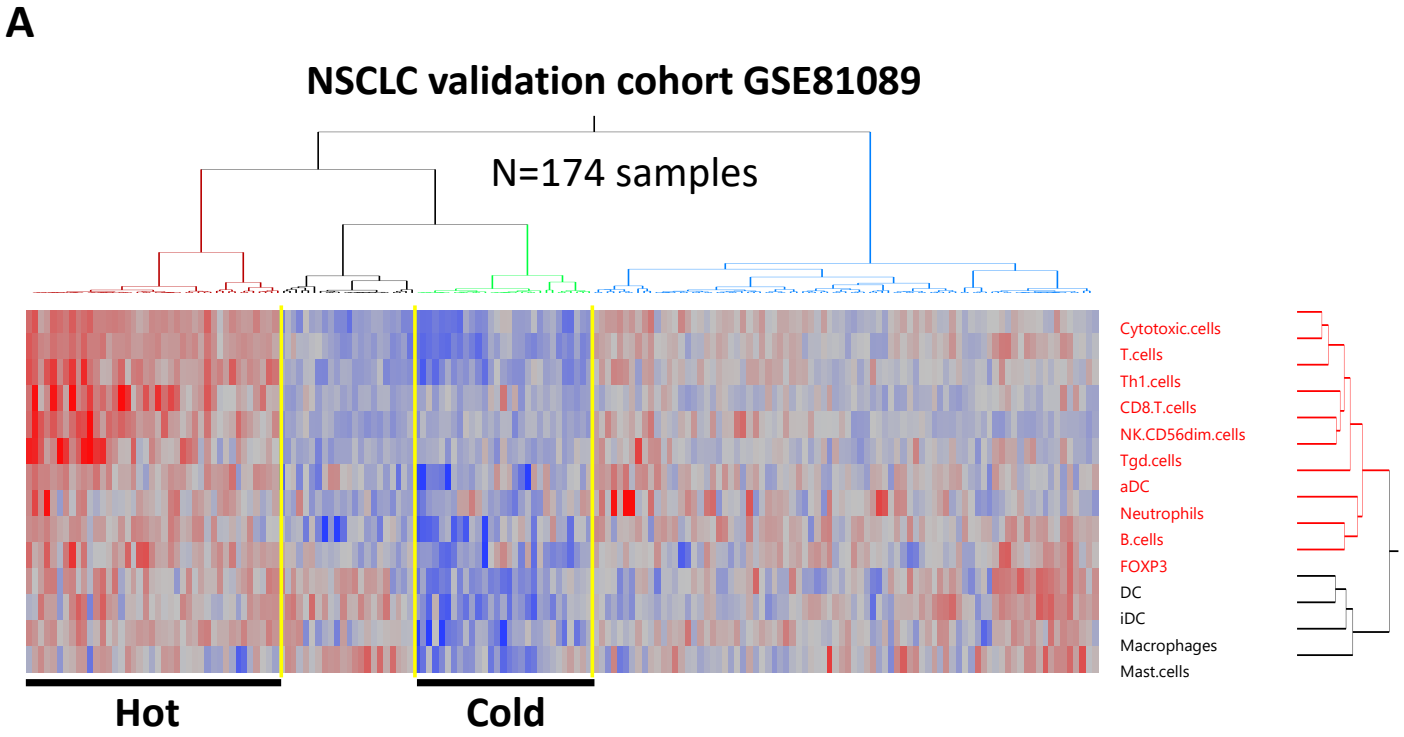


Figure S6. GPX2 expression correlates with immunophenotype across four smoking-related cancers. (A) The average GPX2 expression value for each immune cluster (identified previously in Figure 1) is plotted for all cohorts. (B) Confirmation that GPX2 has a restricted pattern of expression in normal adult 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.



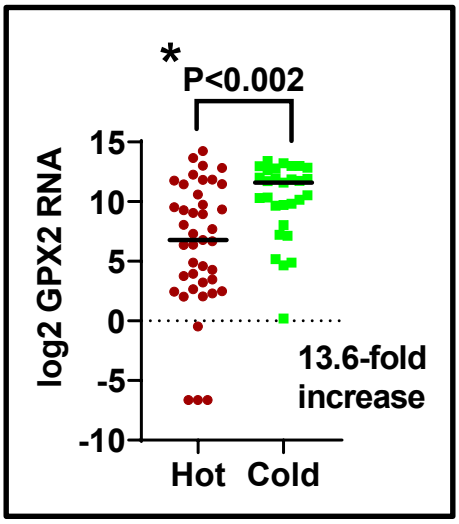
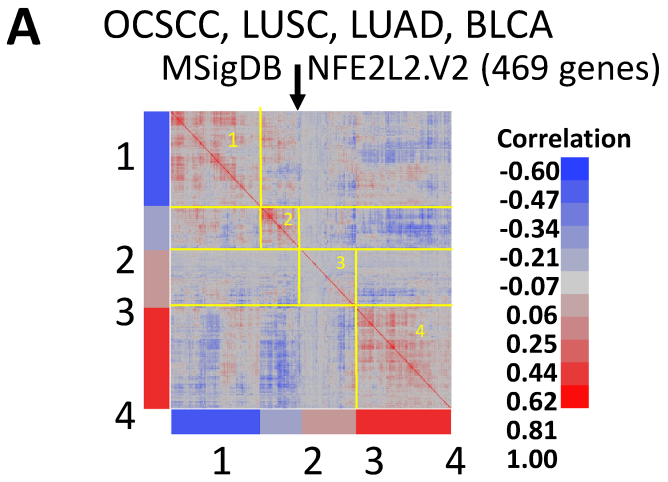


Figure S7. GPX2 overexpression is observed in cold tumors from a TCGA-independent cohort of lung tumors. RNA expression values from a Swedish cohort of lung tumors (GSE81089) consisting of 67 squamous and 107 unspecified histologies, but excluding large cell carcinomas, were converted to FPKM, upper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



Cluster 1 = 138 genes

Go process	Homo sapiens - REFLIST (20595)	Hits	Expected	sign	Enrichmer	P value	FDR
cellular response to xenobiotic stimulus	176	19	1.21	+	15.66	8.51E-17	6.69E-13
oxoacid metabolic process	832	31	5.74	+	5.4	2.11E-14	4.75E-11
cellular response to chemical stimulus	2567	49	17.7	+	2.77	1.33E-11	8.70E-09
lipid metabolic process	1156	30	7.97	+	3.76	3.82E-10	2.14E-07
response to toxic substance	246	13	1.7	+	7.66	2.77E-08	9.69E-06
detoxification	137	9	0.94	+	9.53	7.49E-07	1.61E-04
response to oxidative stress	382	13	2.63	+	4.94	3.38E-06	6.56E-04
NADP metabolic process	31	5	0.21	+	23.39	4.46E-06	8.24E-04
NADPH regeneration	13	4	0.09	+	44.63	4.68E-06	8.55E-04
glutathione metabolic process	64	6	0.44	+	13.6	8.48E-06	1.45E-03
cellular response to oxygen-containing compound	1043	21	7.19	+	2.92	1.18E-05	1.83E-03
cellular detoxification	108	7	0.74	+	9.4	1.44E-05	2.18E-03

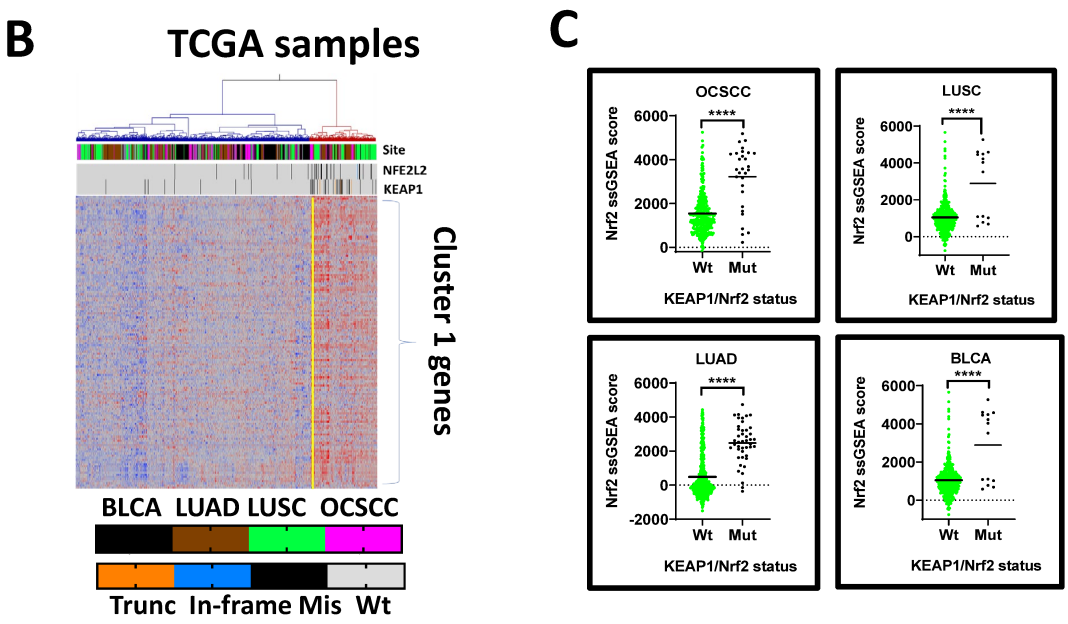


Figure S8. Validation of a gene signature for Nrf2 activation. (A) RNA from expression of 469 genes obtained from the NFE2L2.V2 gene list (i.e., candidate Nrf2 downstream targets) were examined for their cross-correlation among a pooled cohort of TCGA samples encompassing OCSCC, LUSC, BLCA, and LUAD. 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.

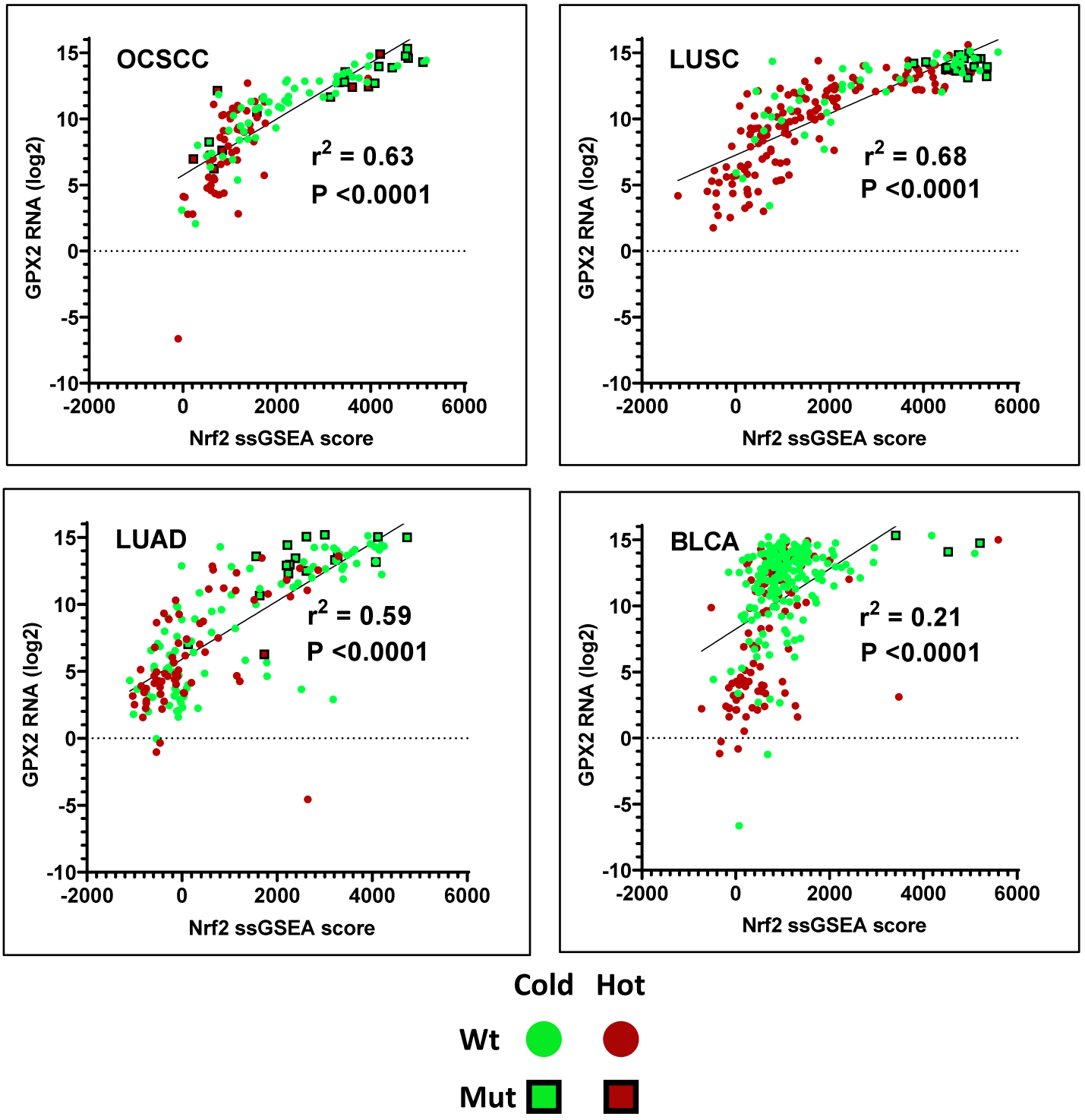


Figure S9. Expression of GPX2 RNA correlates well with Nrf2 activation in three of the smoking-related cancers. The GPX2 RNA expression was plotted versus Nrf2 ssGSEA score for individual tumors, which were visualized according to their immunophenotype (green = cold tumor; red = hot tumor) and mutational 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.

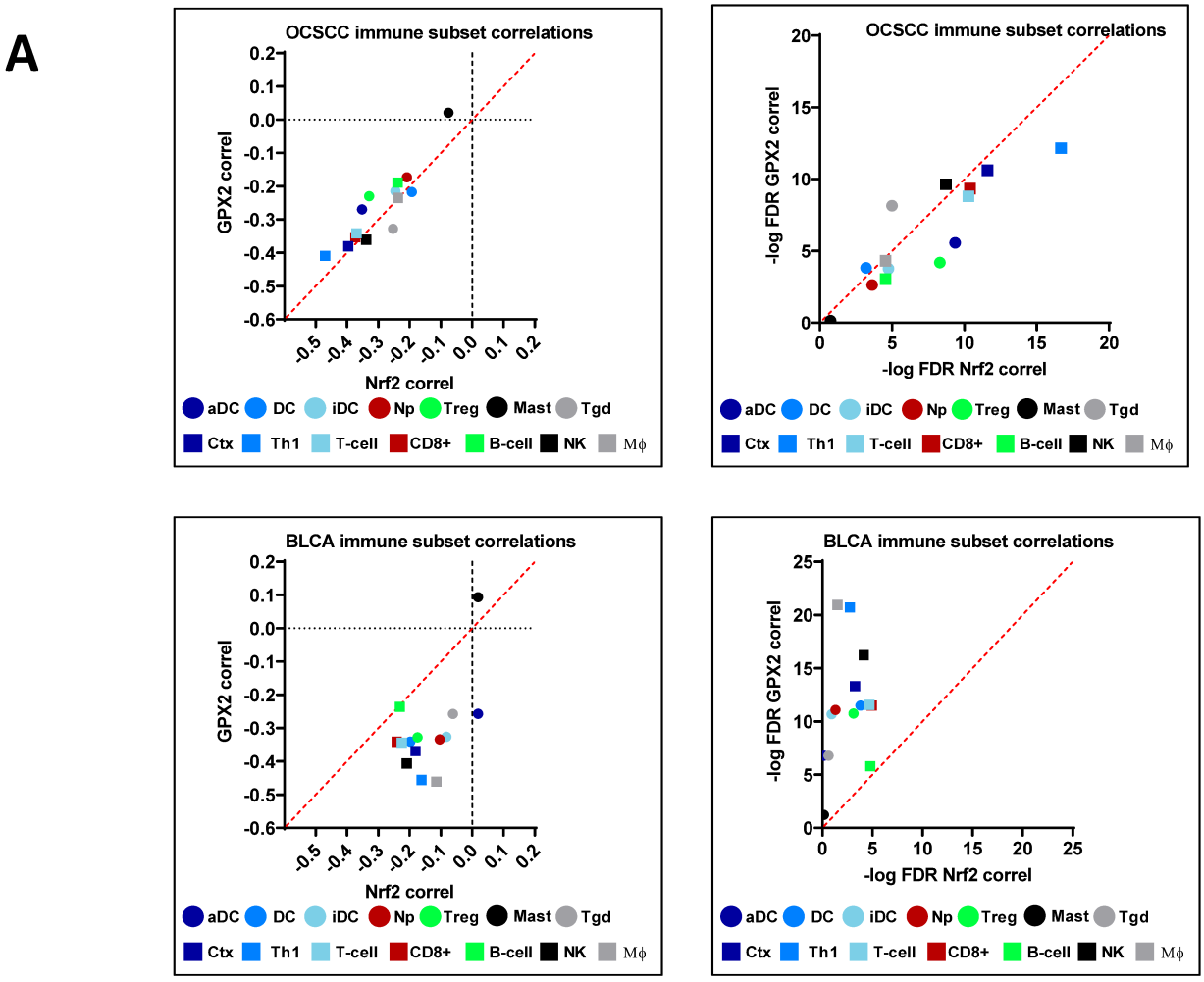


Figure S10. The TIME is more strongly associated with GPX2 than Nrf2 activation in BLCA. (A) The Pearson correlation values between individual leukocyte subtype ssGSEA scores (Supplementary Table S2) and either GPX2 RNA or Nrf2 ssGSEA are plotted for comparison (left panels) or their corresponding FDR values as a measure of significance (right panels). Deviation from the identify 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.

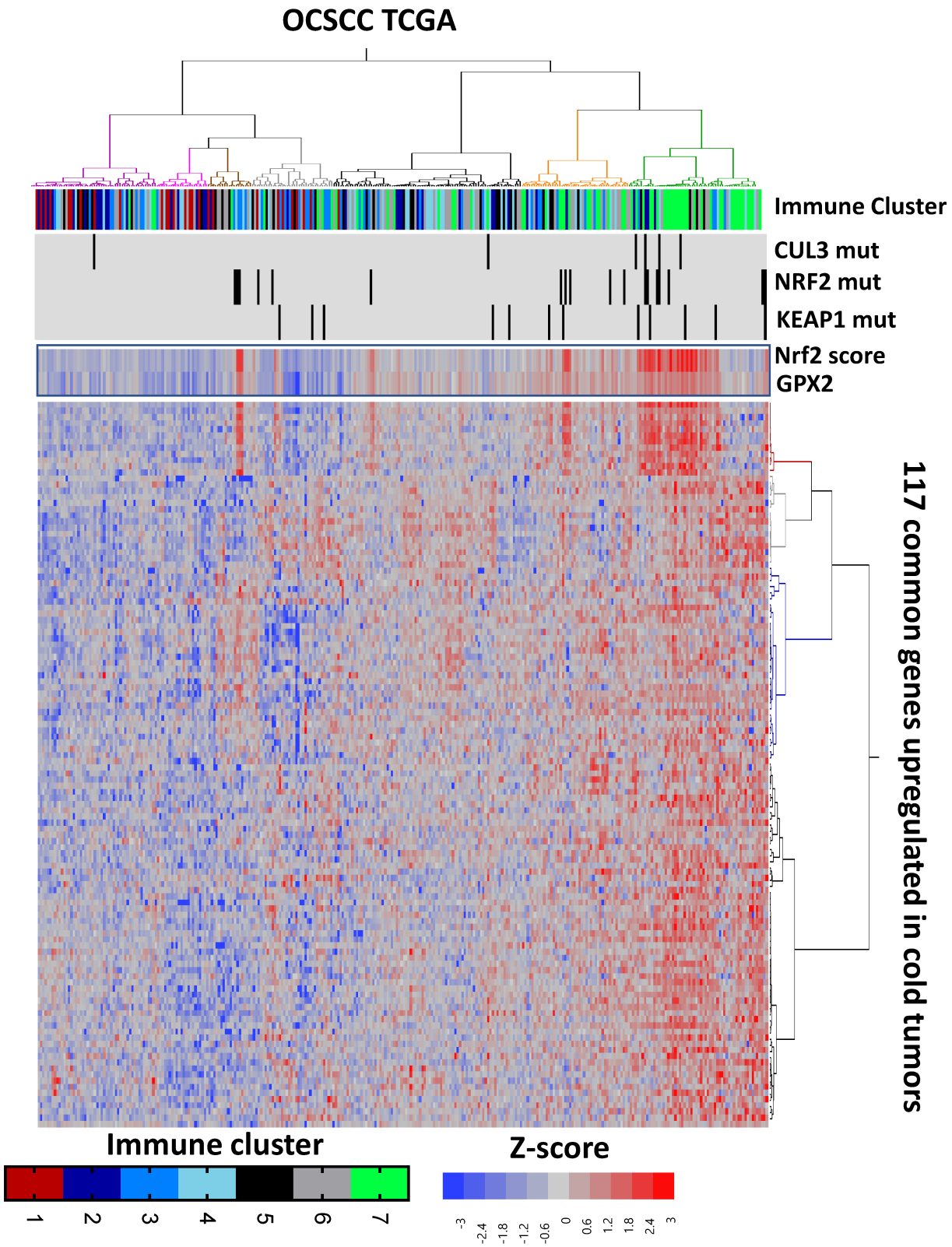
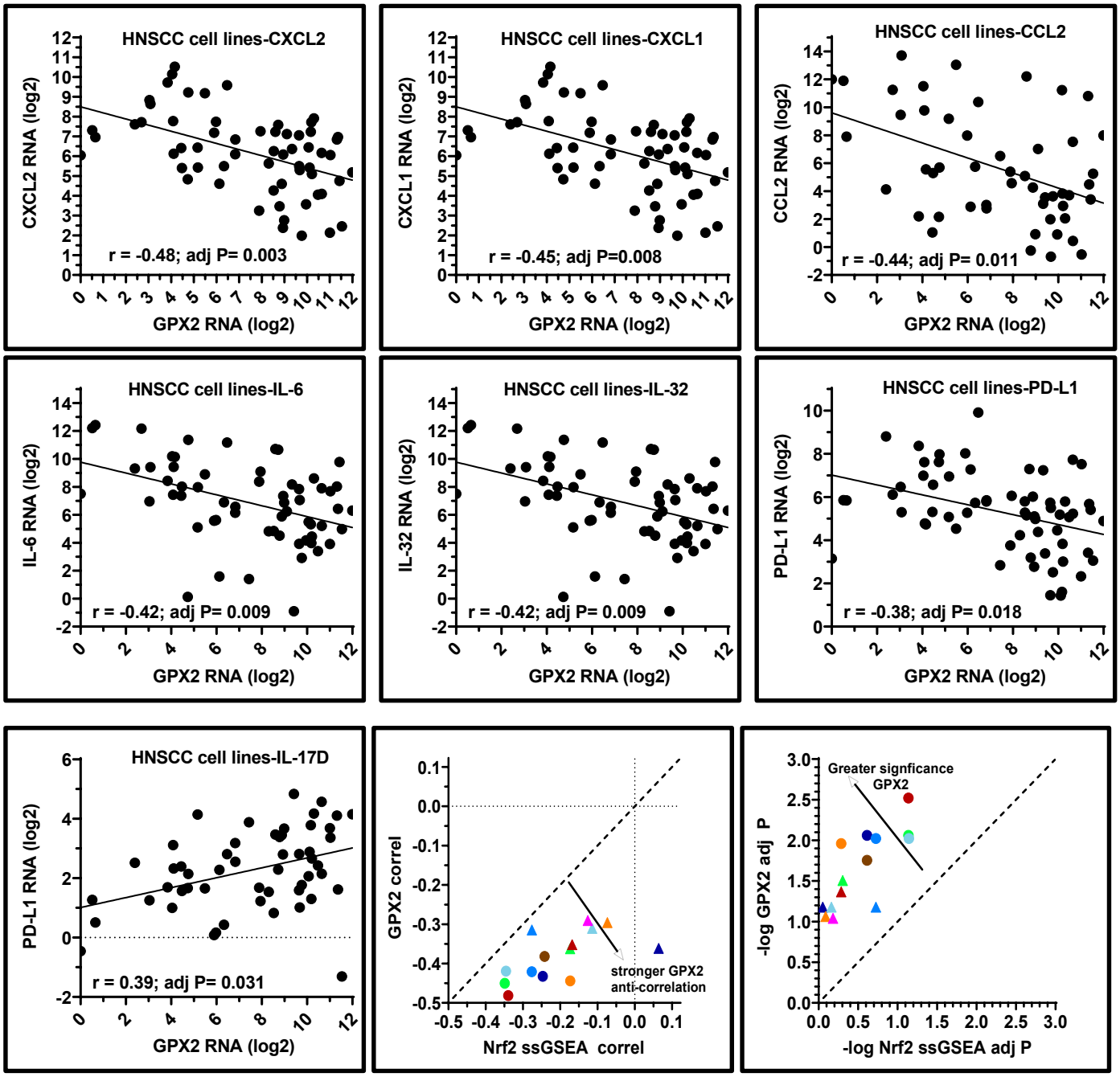
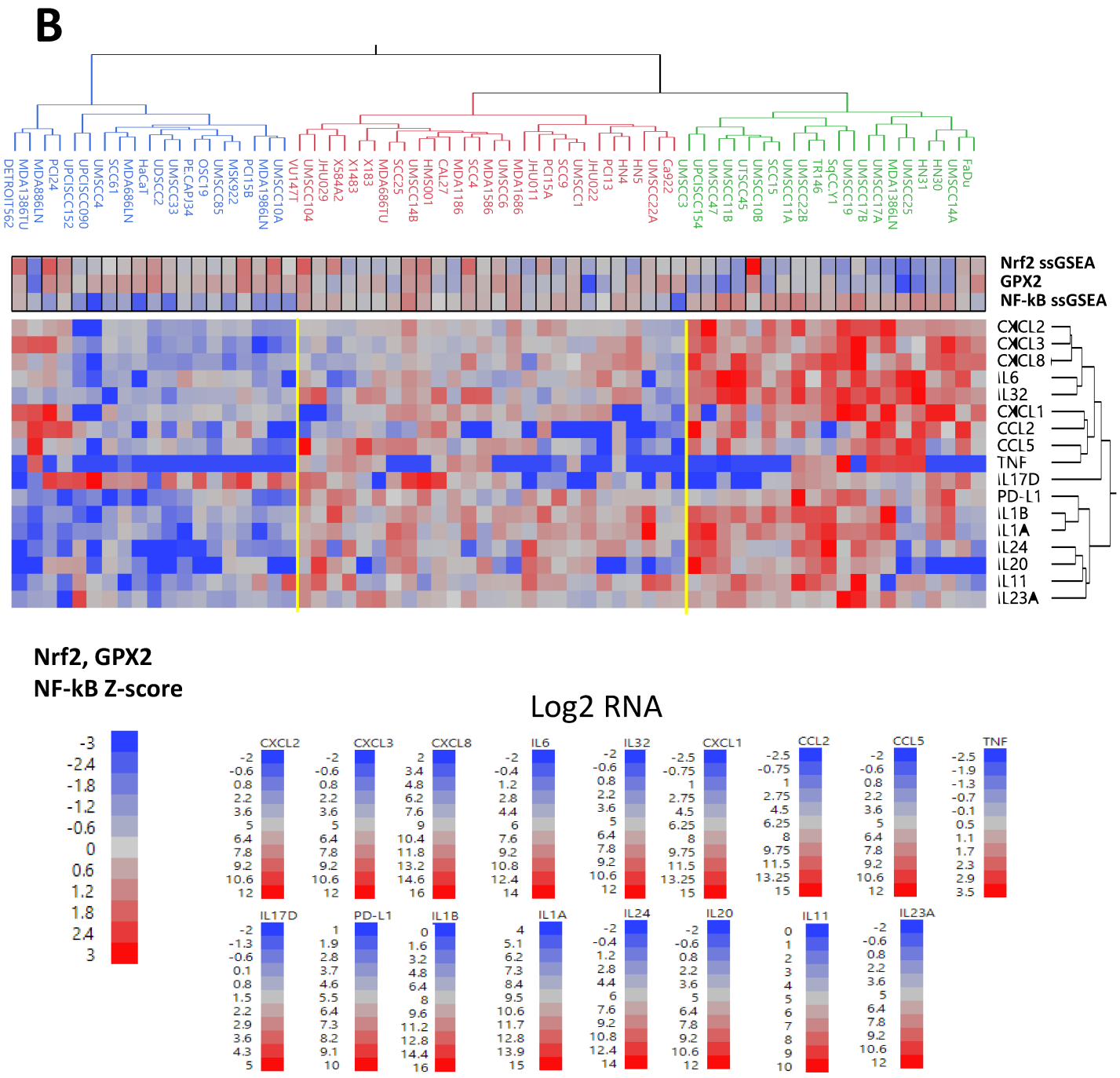


Figure S11. Evidence for Nrf2-dependent and Nrf2-indepdendent regulation of GPX2 and other genes elevated in cold tumors. Expression of the 117 common genes upregulated in cold tumors (including GPX2) was used to cluster TCGA OCSCC samples, which were annotated by immune cluster (i.e. Figure 1), Nrf2 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



Supplementary Figure S12B- See figure legend on next page

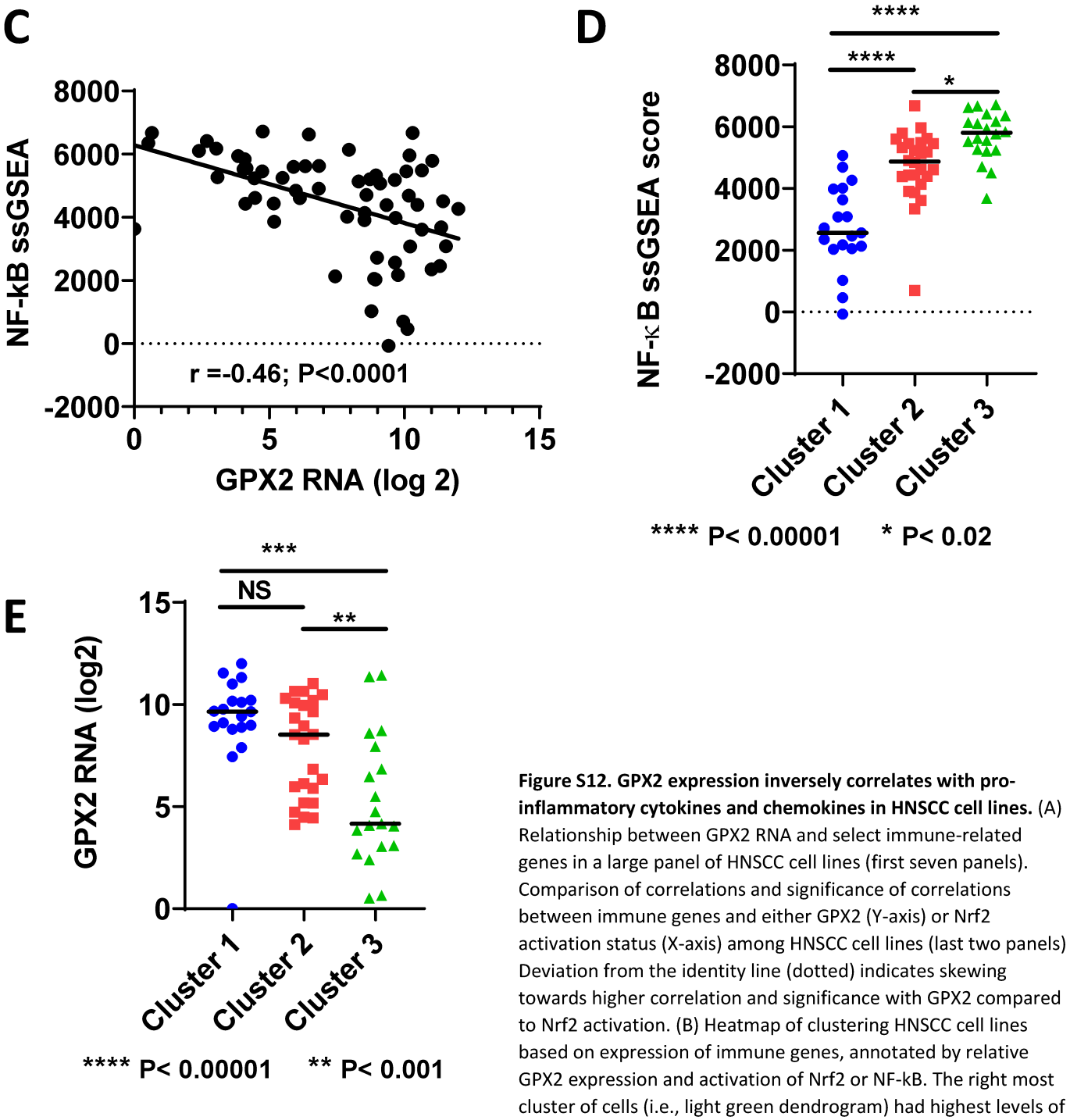


Figure S12. GPX2 expression inversely correlates with pro-inflammatory cytokines and chemokines in HNSCC cell lines. (A) Relationship between GPX2 RNA and select immune-related genes in a large panel of HNSCC cell lines (first seven panels). Comparison of correlations and significance of correlations 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-kB. 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-kB activation, lower GPX2 expression, and lower Nrf2 activation. (C) Significant inverse correlation between GPX2 expression and NF-kB activation in cell lines. (D).NF-kB 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

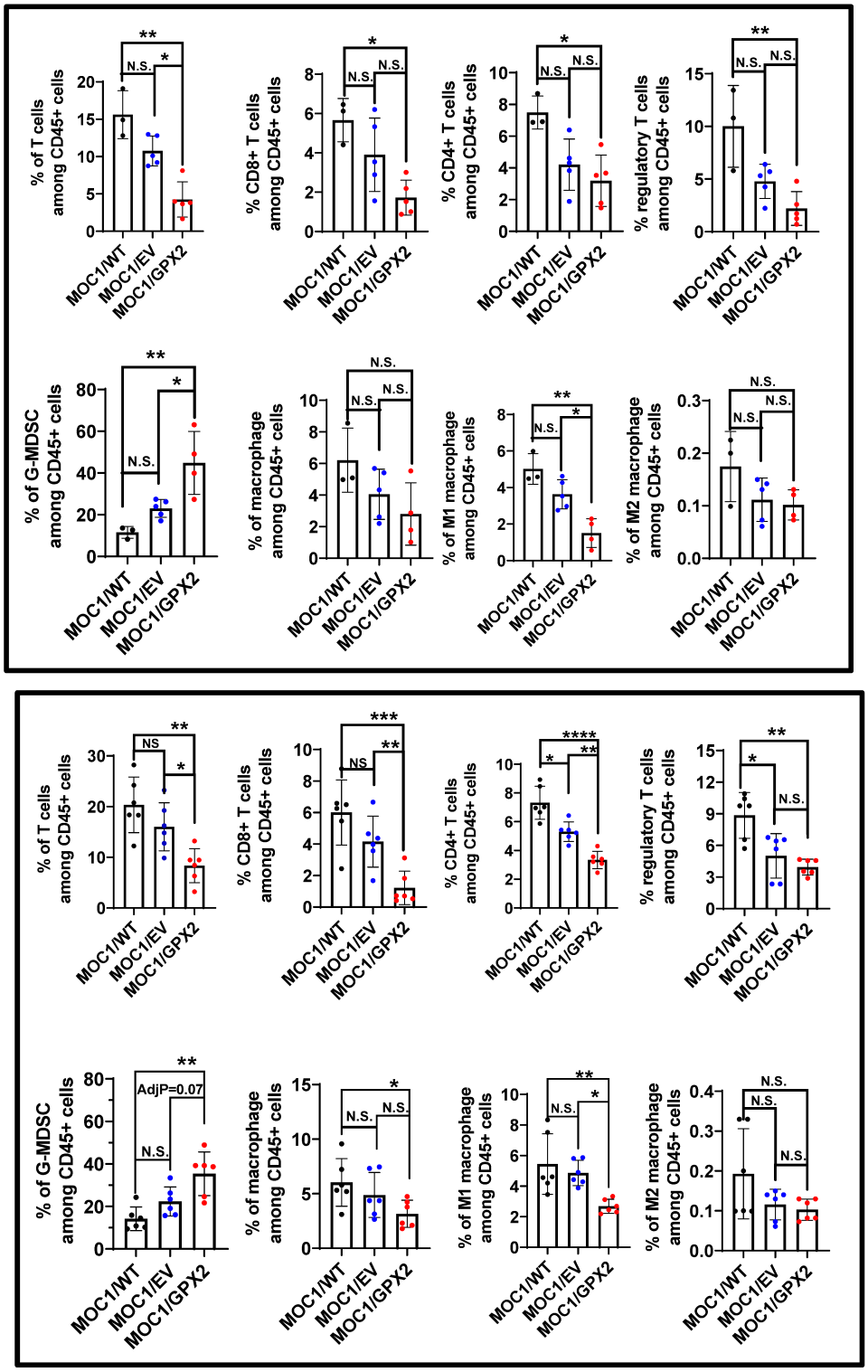


Figure S13. GPX2 drives TIME shifts. Replicate experiments (top and bottom) in which MOC1 tumors (N=3 to 5) generated from parental (WT), empty vector (EV) or GPX2 containing constructs were harvested at the same size and analyzed via flow cytometry. MOC1 parental tumors grew more slowly and had fewer 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 statical analysis.

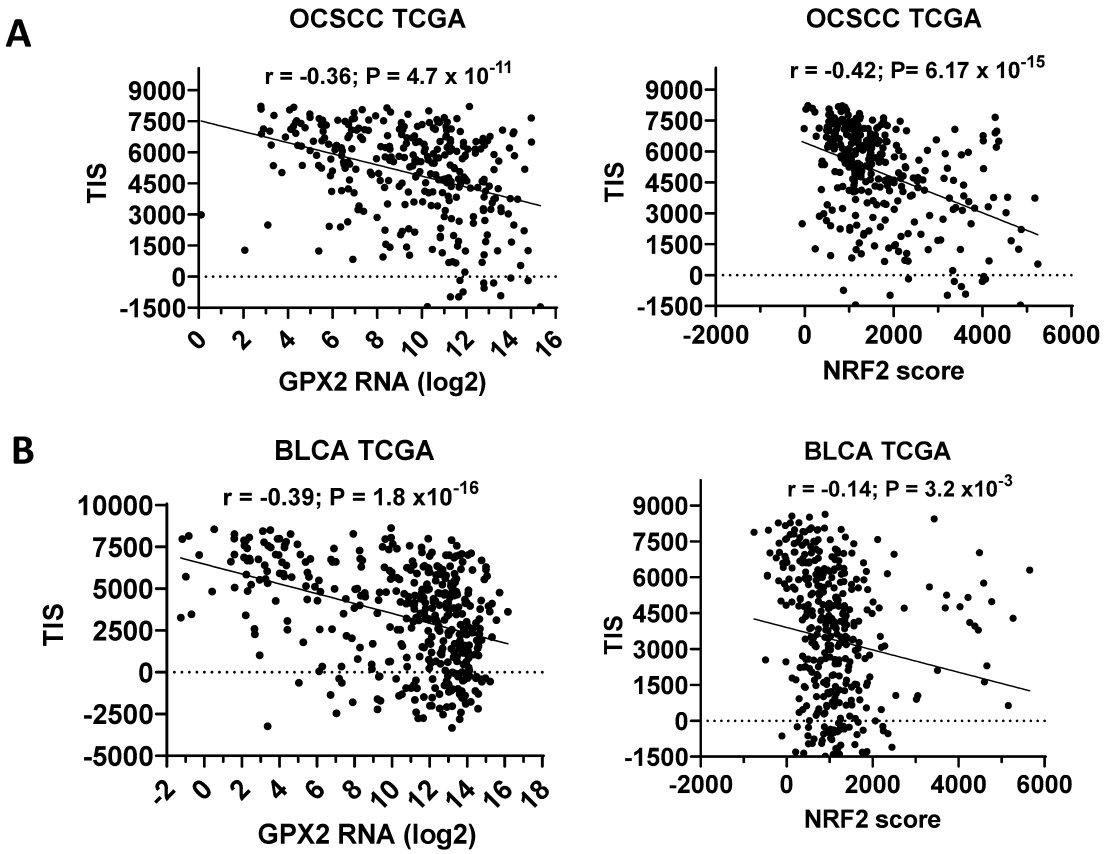


Figure S14. TIS correlates strongly with GPX2 RNA expression. ssGSEA using the 18 gene pan-cancer immune signature published previously by Ayers et al. [37] was used to correlate OCSCC (A) or BLCA (B) TIS score with expression of GPX2 RNA(left panels) or Nrf2 activation (right panels). In BLCA, the correlation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

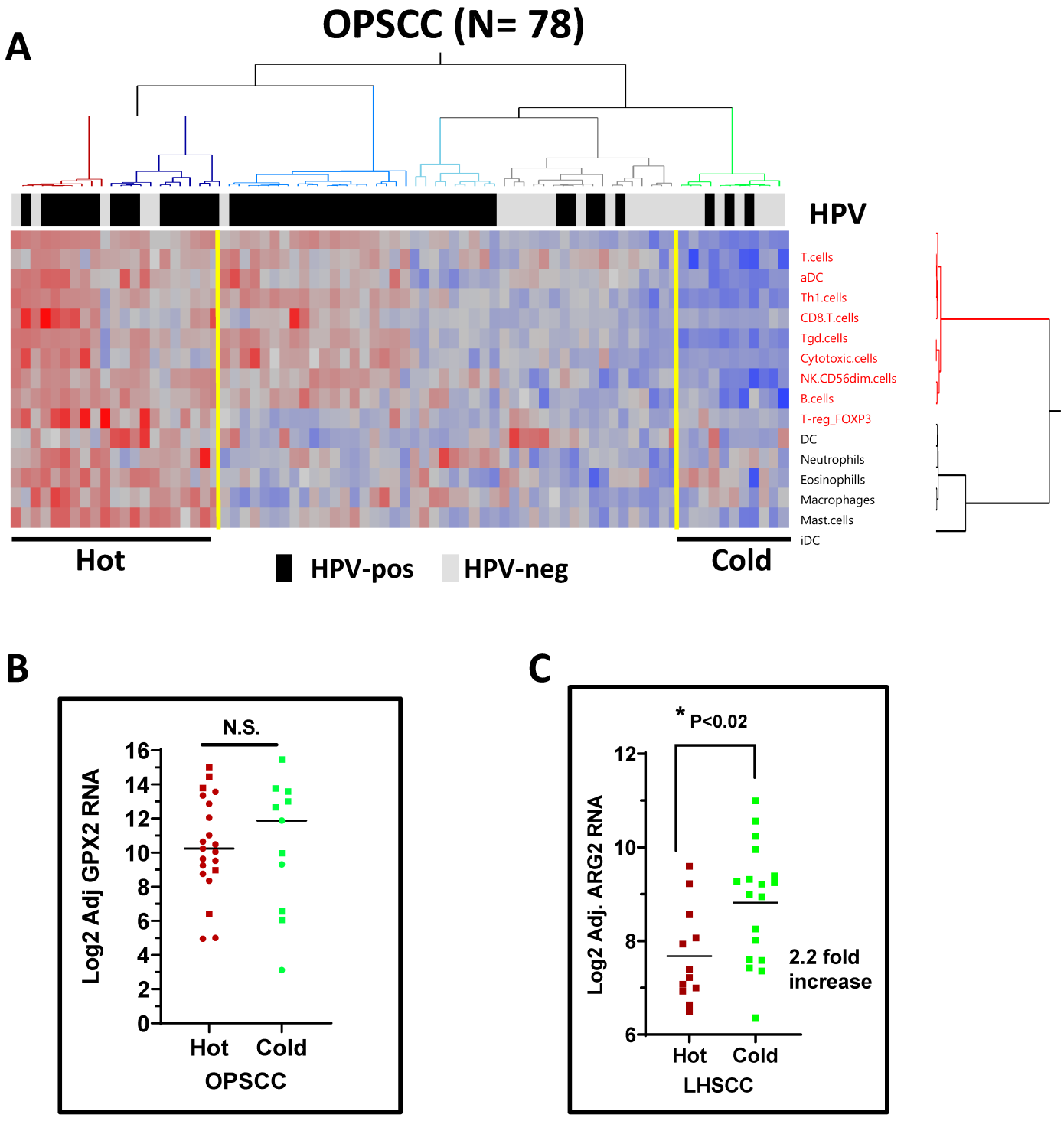


Figure S15. Analysis of cold tumors in OPSCC and LHSCC. (A) Heatmap depicting hot and cold tumors from the TCGA OPSCC samples showing that HPV, frequently present, is strongly correlated with immune infiltration (P =0.02, Fisher's exact test). (B) GPX2 was not significantly correlated with immune phenotype 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.

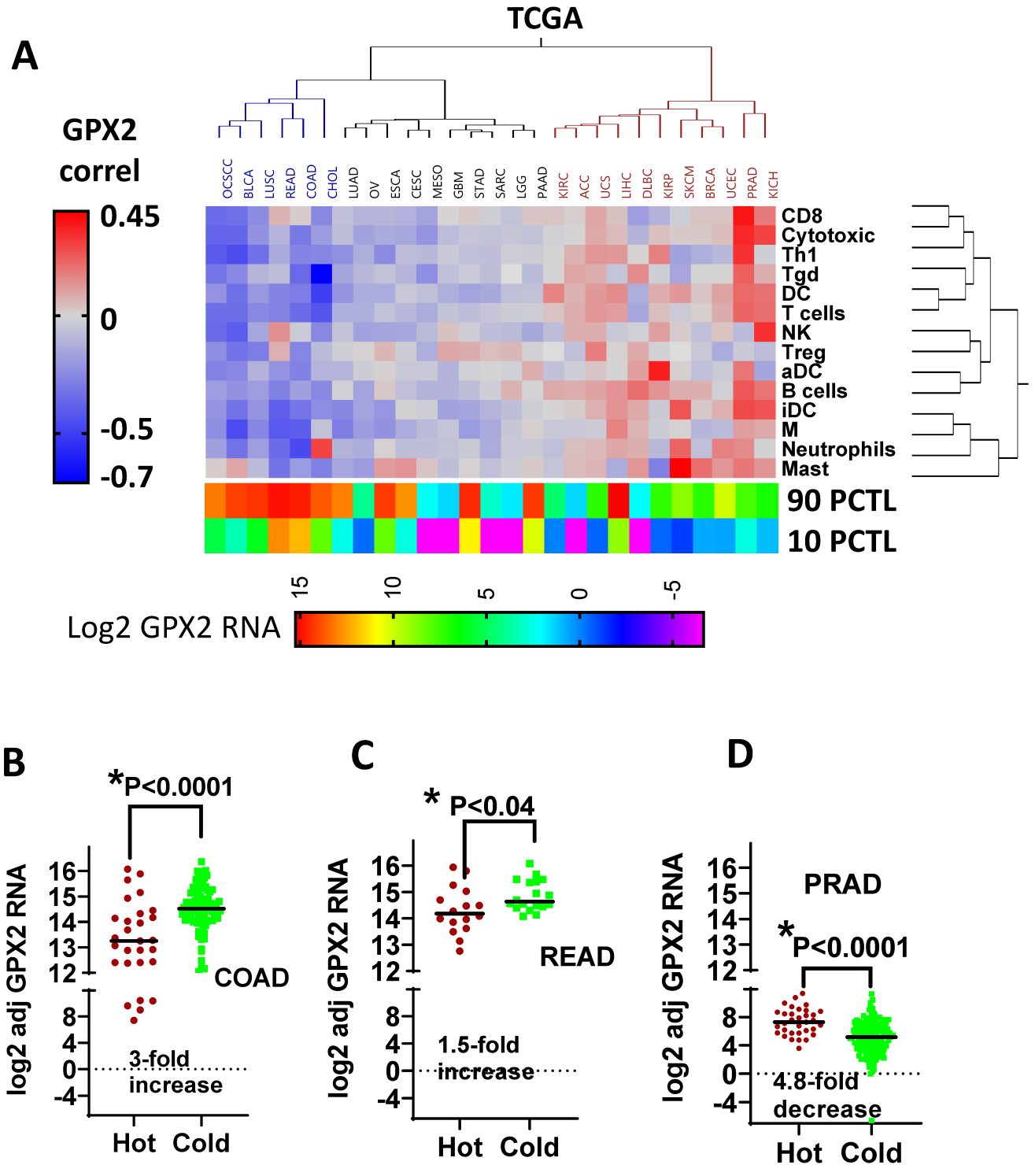


Figure S16. Pan-cancer analysis of GPX2 expression and the TIME. (A) Correlation between GPX2 RNA and degree of infiltration by leukocyte subsets (i.e., ssGSEA scores or log2 FOXP3 or Tregs) across smoking related cancers and 23 additional tumor types studied by the TCGA. Correlation values were used for 2 ways Ward's clustering. The relative range of GPX2 expression for each cohort is represented by the10th 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