

Proteogenomic and metabolomic characterization of human glioblastoma - Wang et al. 2021

- genomic data analysis
 - copy number variant calling
 - somatic variant calling
 - germline variant calling
 - TERT promoter mutation calling
 - structural variant calling
 - DNA methylation
 - classification of MGMT promoter DNA methylation status
 - telomere length quantification and genotyping (estimate the telomere length using WXS and WGS tumor and blood normal paired BAMs)
- RNA analysis
 - RNA quantification
 - RNA fusion detection
 - miRNA quantification
 - circular RNA prediction and quantification
- snRNA-seq analysis
- MS data
 - quantification of phosphopeptides
 - quantification of acetylated peptides
 - metabolome and lipidome
- other analyses
 - ancestry prediction using SNPs from 1000 genomes project
 - multi-omics subtyping using NMF
 - expression-based TCGA subtyping
 - unsupervised clustering of DNA methylation
 - MolecularNeuroPathology (MNP) DNA Methylation Classification of Central Nervous System (CNS) Tumors
 - unsupervised clustering of miRNA expression Unsupervised
 - unsupervised clustering of individual data type
 - determination of stemness score
 - multi-omics cis association analysis using iProFun (integrate somatic mutation, CNV, DNA methylation, RNA, protein, phosphorylation (phospho) and acetylation (acetyl) levels via iProFun (Song et al., 2019) to investigate the functional impacts of DNA alterations in GBM.)
 - mutation impact on the RNA, proteome, phosphoproteome, lipidome and metabolome
 - protein and RNA marker identification for multi-omics mixed subtype
 - kinase-substrate pairs regression analysis
 - differential proteomic, phosphosite, metabolome and lipidome analysis
 - copy number impact on transcriptome and proteome
 - cell type enrichment deconvolution using gene expression
 - immune clustering using cell type enrichment scores
 - cell type enrichment deconvolution using protein abundance
 - deep learning histopathology image analysis
 - gene set enrichment analysis
 - histone protein and acetylation calculation

- histone acetylation association with HATs and HDACs
- pathway enrichment analysis along histones H2B and H3/H4 acetylation axes
- causative pathway interaction discovery using CausalPath
- L1000 and P100 drug connectivity analysis

CausalPath

Application of CausalPath (Babur et al., 2018) to the protein and phosphoprotein expression data (Figure S7A, Table S5) showed upregulation of the hypoxia pathway in mesenchymal tumors, evidenced by significant activation of multiple HIF-1 downstream targets (network permutation $p = 0.0012$).

To discover the causative pathway interactions in our proteomic and phosphoproteomic data, we took the **normalized expression of protein** with < 10% missing values **and phosphoprotein** with < 25% missing values across all tumor and normal samples as the input to CausalPath (commit 7c5b934). We ran CausalPath in the mode that **tests the mean values between test and control groups** (value-transformation = significant-change-of-mean), where the test group being the tumors of one subtype and control group being the rest of the tumors. The pathway interaction discovery data source was Pathway Commons v9 (built-in- network-resource-selection = PC). Additionally, we enabled the causal reasoning if all the downstream targets of a gene were active or inactive (calculate-network-significance = true, use-network-significance-for-causal-reasoning = true, permutations- for-significance = 10000). The causative interactions with FDR < 0.05 were extracted and visualized (fdr-threshold-for-data-significance = 0.05 phosphoprotein, fdr-threshold-for-data-significance = 0.05 protein, fdr-threshold-for-network-significance = 0.05). Full result tables were available in Table S5.

A proteogenomic portrait of lung squamous cell carcinoma - Satpathy et al. 2021

- genomic data analysis
 - copy number calling
 - somatic variant calling
 - germline variant calling
 - GISTIC and mutSig analysis
- RNA data analysis
 - RNAseq and miRNAseq quantification and analysis
 - isoform specific RNA quantification
 - DNA methylation data preprocessing
- proteomics data analysis
- proteogenomic analysis
 - unsupervised multi-omic clustering using NMF
 - integrative analysis with Stewart et al. (describe a way to integrate their protein data with protein data from another dataset !)
 - RNA subtyping
 - chromosomal instability index (CIN)
 - fusion detection and analysis
 - mRNA and Protein correlation

- CNA-driven cis and trans effects
- CMAP analysis
- LINCS analysis
- defining cancer-associated genes
- CpG island methylator phenotype
- iProFun based cis association analysis (iProFun = an integrative analysis tool to identify multi-omic molecular quantitative traits (QT) perturbed by DNA-level variations)
- EMT-derived cluster and fibroblast enrichment
- differential marker analysis (a Wilcoxon signed rank test was performed on TMT-based global proteomic data between tumor and matched normal samples to determine differential abundance of proteins between tumor and NAT samples)
- survival analyses
- immune cluster identification based on cell type composition
- TCGA pan-cancer immune subtyping
- ranking tumors by inferred activity of IFN-g pathway
- estimation of tumor purity, stromal and immune scores
- differentially expressed genes and pathway analysis
- independent component analysis (correlation between the extracted signatures and known clinical characteristics were examined by regressing the corresponding mixing scores for all members of a component cluster against 64 sample annotations to obtain within-cluster average of log₁₀ p values)
- mutation-based cis- and trans-effects (we examined the cis- and trans-effects of 22 genes with somatic mutations that were significant in a previous large-scale TCGA LSCC study (Bailey et al., 2018) on the RNA, proteome, and phosphoproteome of known interactome DBs including Omnipath, Phosphositeplus, DEPOD, Signor, and CORUM)
- germline quantitative trait loci (QTL) analysis (to identify germline genetic variants that explain variation in tumor gene (eQTL) and protein (pQTL) expression, we utilized the gold- standard mapping pipeline)
- pathway projection using ssGSEA
- phosphorylation-driven signature analysis (we performed phosphosite-specific signature enrichment analysis (PTM-SEA) (Krug et al., 2019) to identify dysregulated phosphorylation-driven pathways)
- cluster and pathway analysis of significantly modulated K-GG sites in tumors
- DepMap genetic dependency and drug response analysis
- CausalPath analysis (CausalPath (Babur et al., 2018) searches for known biological mechanisms that can explain correlated proteomic changes in terms of causal hypotheses)
- variant peptide identification We used NeoFlow (<https://github.com/bzhanglab/neoflow>) for neoantigen prediction
- cancer/testis antigen prediction (cancer/testis (CT) antigens were downloaded from the CTdatabase (Almeida et al., 2009))
- PROGENy scores (PROGENy (Schubert et al., 2018) was used to generate activity scores for EGFR based on RNA expression data)

CausalPath

CausalPath (Babur et al., 2018) **searches for known biological mechanisms that can explain correlated proteomic changes in terms of causal hypotheses.** We set CausalPath parameters to **compare tumors and NATs** [NAT=normal adjacent tissue] with a paired t test, used 0.1 as FDR threshold **for proteomic change significance and network significance**, and detected **5917 potential causal relations between proteins.** We repeated the **same analysis for each NMF subtype separately** and identified 4378 (basal-inclusive), 5334 (classical), 3048 (EMT-enriched), 3744 (inflamed-secretory), and 4332 (proliferative-primitive) relations. We used these CausalPath

network results in the preparation of Figure 7C, **identifying potential upstream regulators of oncogenic phosphoproteomic changes**. Here an **oncogenic phosphoproteomic change** can be any of the following 4 events: increase of activating phosphorylation of an oncoprotein, decrease of inactivating phosphorylation of an oncoprotein, decrease of activating phosphorylation of a tumor suppressor protein, and increase of inactivating phosphorylation of a tumor suppressor protein. We used the OncoKB database for oncoprotein and tumor suppressor classification (excluded proteins that have both annotations), and used PhosphoSitePlus and Signor databases for the activating/inhibiting classification of phosphorylation sites. In the phosphorylation regulation networks, we included only the targetable regulators (activated proteins) and excluded the untargetable regulators (inactivated proteins).

Figure 7: For identifying if top protein biomarkers (502 proteins significantly overexpressed ($\log_2(\text{FC}) > 2$, $\text{FDR} < 0.01$) in tumors relative to their matched NATs, most with coherent overexpression in multi-omic analysis) also conferred altered dependencies in LSCC cell lines, we leveraged DepMap genetic dependency dataset (CRISPR Avana Public 20Q3) that contained 18119 genes and 789 cell lines (<https://depmap.org/portal/download/> file: Achilles_gene_effect.csv). Only 16 cell lines were classified as LSCC (Cell Line Sample Info.csv). Median dependencies were calculated and for every gene and density plot in Figure 7C shows dependency score of the 502 genes corresponding to 502 protein biomarkers relative to other genes.

Figure 7(C): Genetic dependencies of 502 proteins ($\log_2 \text{FC} > 2$, $\text{FDR} < 0.01$ and NAs $< 50\%$) in LSCC cell lines ($n = 16$) profiled as part of the Achilles Dependency-Map project

[...] we leveraged the TCGA LSCC dataset (Hammerman et al., 2012) to examine the association of these 502 candidate tumor biomarkers with overall survival (OS) or disease-free survival (DFS). Expression of four of the most highly differential genes showed significant association with poor OS and another 15 genes with poor DFS (Figure 7B; Figure S7C). Furthermore, knockdown of these tumor biomarker proteins reduced fitness across 16 LSCC cell lines (<https://depmap.org/>), suggesting critical roles in key cellular transformation and proliferation processes (Figure 7C; Table S7)