

Method S1. Primary organization and functionalities of the LinkedOmicsKB web pages.

The LinkedOmicsKB web portal assembles precomputed analysis results into ~40,000 gene-, protein-, mutation-, and phenotype-centric web pages. Here we describe the primary organization and functionalities of these web pages.

A gene page includes six sections. The basic information section shows the gene description from the RefSeq database¹ and functional annotations compiled from multiple sources, such as cancer driver, kinase, and drug target annotations. The protein product section lists all protein products of the gene, with links to protein pages (see below) of systematically selected primary and secondary proteins (**Methods**). The tumor vs normal comparison section presents differential analysis results at mRNA and protein levels, respectively. The phenotype and mutation association section begins with a Manhattan plot summarizing associations between the gene and all clinical phenotypes, molecular phenotypes, and somatic mutations across all cancer types at copy number, mRNA, and protein levels, respectively (**Fig. 2a**). This plot enables quick identification of cancer types and omics data types with interesting associations, as well as interactive examination of the top, highly significant associations. Detailed association results are presented in a sortable, searchable, filterable, and expandable table that can be switched between protein, mRNA, and copy number views (**Fig. 2b**). The table was designed to hold four dimensions (4D) of information, with columns corresponding to individual cancer types or pan-cancer, primary rows corresponding to associations between a phenotype/mutation and the omics data type of the primary view, expandable rows displaying associations at other omics levels for multi-omics comparison, and pop-up windows with appropriate statistical plots supporting the signed p-values displayed in the table, such as scatter plot, box plot or Kaplan-Meier plot. Phenotypes and mutations in the table are linked to corresponding phenotype and mutation pages (see below), and a heatmap summarizing pan-cancer, multi-omics association results is also available for each phenotype or mutation (**Fig. 2c**). The *cis*-association section depicts pairwise associations between protein, mRNA, copy number, and methylation levels of the gene within each cancer type in a correlogram (**Fig. 2d**), in which each circle can be clicked to show the corresponding scatter plot. The *cis*-association between different omics measurements could hint at interesting regulatory mechanisms². The *trans*-association section displays association results between the gene and all other protein

coding genes at mRNA and protein levels, respectively, in a 4-D table similar to the phenotype association table. In addition to providing an easy way to navigate through the top associated genes, the complete *trans*-association results for individual data types in a selected cancer type or the pan-cancer results can be forwarded to WebGestalt³ with a single click for pathway and gene set enrichment analysis (**Fig. 2h**). This provides a means for gene function inference based on the guilt-by-association hypothesis, which is particularly compelling when applied to protein association data^{4,5}.

The protein page includes information for all experimentally identified phosphosites from the proteins in CPTAC studies, divided into seven sections. The first section uses a zoomable lollipop plot and a 3D protein structure viewer⁶ to visualize all identified phosphosites in the context of protein sequence, domains, and structure (**Fig. 2f**). The site detection section provides further information on the detection frequencies within and across all cancer types in an interactive table. The sections for tumor vs normal comparison and phenotype and mutation association are similar to those in the gene pages, except for using phosphosites, rather than genes, as the units for analysis. The *cis*-association section visualizes associations between individual phosphosites and measurements at protein, mRNA, copy number and methylation levels, respectively, in a correlogram, which is useful for prioritizing phosphosites that are regulated independent of protein abundance (**Fig. 2e**). The last two sections use interactive tables to display association results between phosphosites and all kinases and all phosphatases, respectively, which may help prioritize direct or indirect upstream regulators of the phosphosites.

Users can also browse precomputed results in the phenotype or mutation pages. These pages first introduce basic information of the phenotype or mutation, such as the method used to quantify a molecular phenotype. Next, association results of all genes and phosphosites with the phenotype or mutation are presented. Four two-sided bar charts summarize the numbers of significant associations ($p < 1.0e-6$) in each cancer type at copy number, mRNA, protein, and phosphosite levels, respectively. Moreover, one scatter plot comparing mRNA and protein associations and another comparing protein and phosphosite associations (**Fig. 2g**) help prioritize protein and phosphosite-specific associations, respectively. Detailed association results are displayed in two interactive tables for copy number/mRNA/protein and phosphosites,

respectively. Similar to these pages, a specific page is dedicated to tumor vs normal comparison and can be directly accessed from the homepage.

References:

1. O’Leary, N.A., Wright, M.W., Brister, J.R., Ciuffo, S., Haddad, D., McVeigh, R., Rajput, B., Robbertse, B., Smith-White, B., Ako-Adjei, D., et al. (2016). Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res.* *44*, D733–D745.
2. Buccitelli, C., and Selbach, M. (2020). mRNAs, proteins and the emerging principles of gene expression control. *Nat. Rev. Genet.* *21*, 630–644.
3. Liao, Y., Wang, J., Jaehnig, E.J., Shi, Z., and Zhang, B. (2019). WebGestalt 2019: gene set analysis toolkit with revamped UIs and APIs. *Nucleic Acids Research* *47*, W199–W205. [10.1093/nar/gkz401](https://doi.org/10.1093/nar/gkz401).
4. Wang, J., Ma, Z., Carr, S.A., Mertins, P., Zhang, H., Zhang, Z., Chan, D.W., Ellis, M.J.C., Townsend, R.R., Smith, R.D., et al. (2017). Proteome Profiling Outperforms Transcriptome Profiling for Coexpression Based Gene Function Prediction. *Mol. Cell. Proteomics* *16*, 121–134.
5. Kustatscher, G., Collins, T., Gingras, A.-C., Guo, T., Hermjakob, H., Ideker, T., Lilley, K.S., Lundberg, E., Marcotte, E.M., Ralser, M., et al. (2022). Understudied proteins: opportunities and challenges for functional proteomics. *Nat. Methods* *19*, 774–779.
6. Rose, A.S., Bradley, A.R., Valasatava, Y., Duarte, J.M., Prlic, A., and Rose, P.W. (2018). NGL viewer: web-based molecular graphics for large complexes. *Bioinformatics* *34*, 3755–3758.

Table S1. Summary of tumor and adjacent normal tissue numbers.

Cancer Cohort	Tumor Samples	Adjacent Normal Tissue Samples
Breast invasive carcinoma (BRCA)	122	
Clear cell renal cell carcinoma (CCRCC)	103	80
Colon adenocarcinoma (COAD)	110	100
Glioblastoma (GBM)	99	
Head and neck squamous cell carcinoma (HNSCC)	108	62
Lung adenocarcinoma (LUAD)	110	101
Lung squamous cell carcinoma (LSCC)	108	99
Ovarian serous cystadenocarcinoma (OV)	83	20
Pancreatic ductal adenocarcinoma (PDAC)	105	44
Uterine corpus endometrial carcinoma (UCEC)	95	18
Total	1043	524

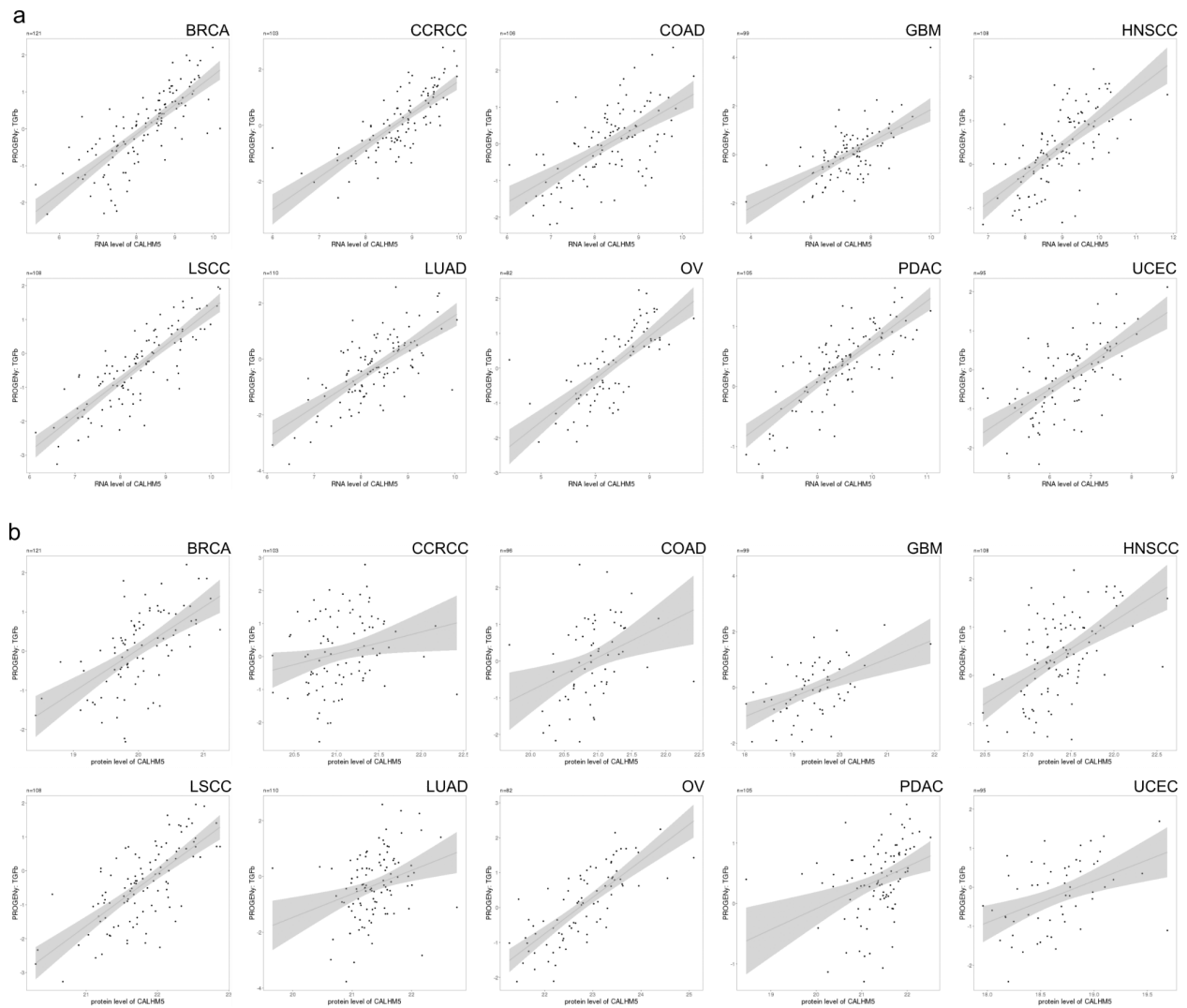


Figure S1. Association of CALHM5 and TGFbeta signaling. (a) Scatter plots of associations between CALHM5 mRNA abundance and TGFbeta signaling signature score in all cancer cohorts. (b) Scatter plots of associations between CALHM5 protein abundance and TGFbeta signaling signature score in all cancer cohorts.

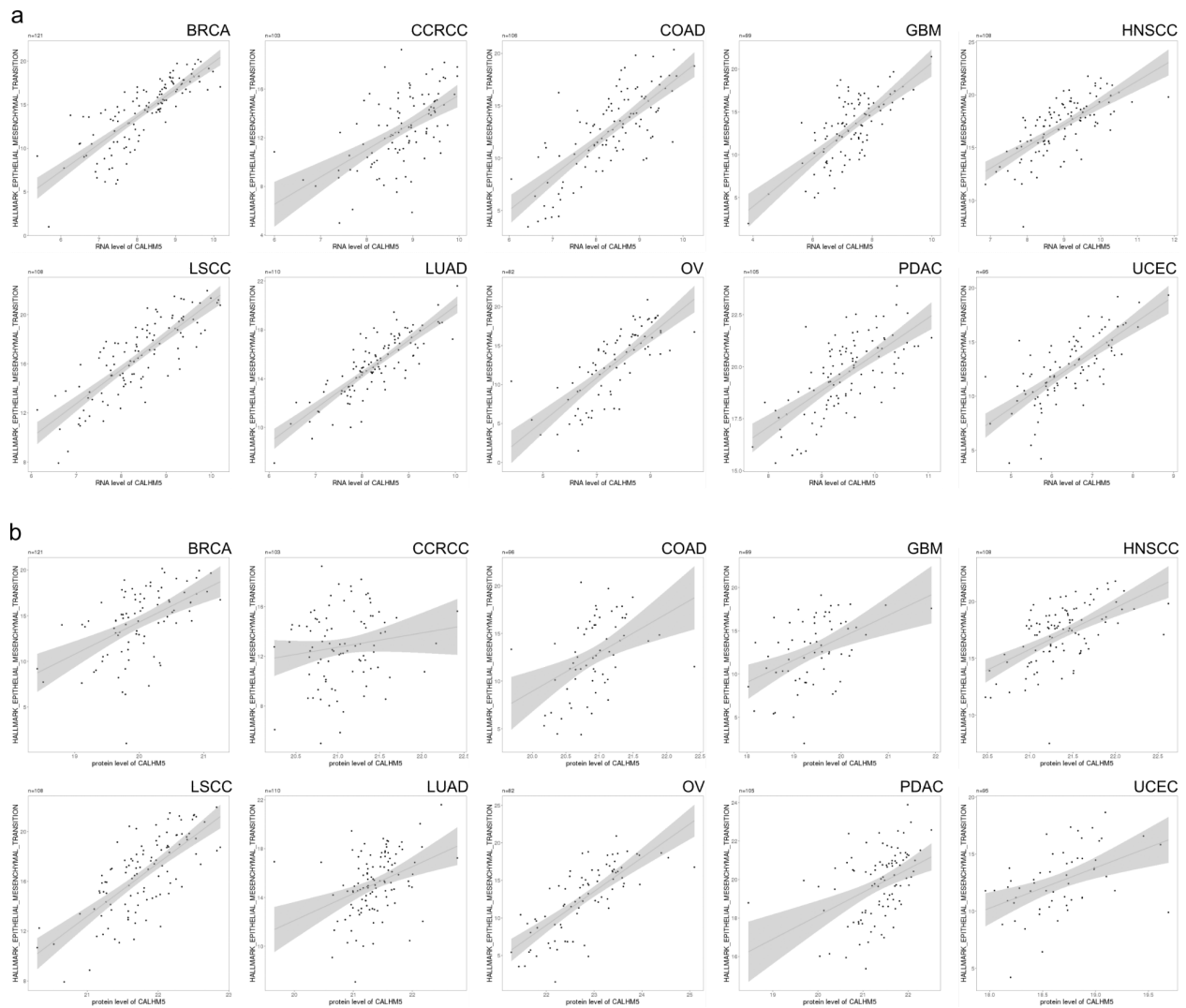


Figure S2. Association of CALHM5 and epithelial-mesenchymal transition. (a) Scatter plots of associations between CALHM5 mRNA abundance and epithelial-mesenchymal transition signature score in all cancer cohorts. (b) Scatter plots of associations between CALHM5 protein abundance and epithelial-mesenchymal transition signature score in all cancer cohorts.

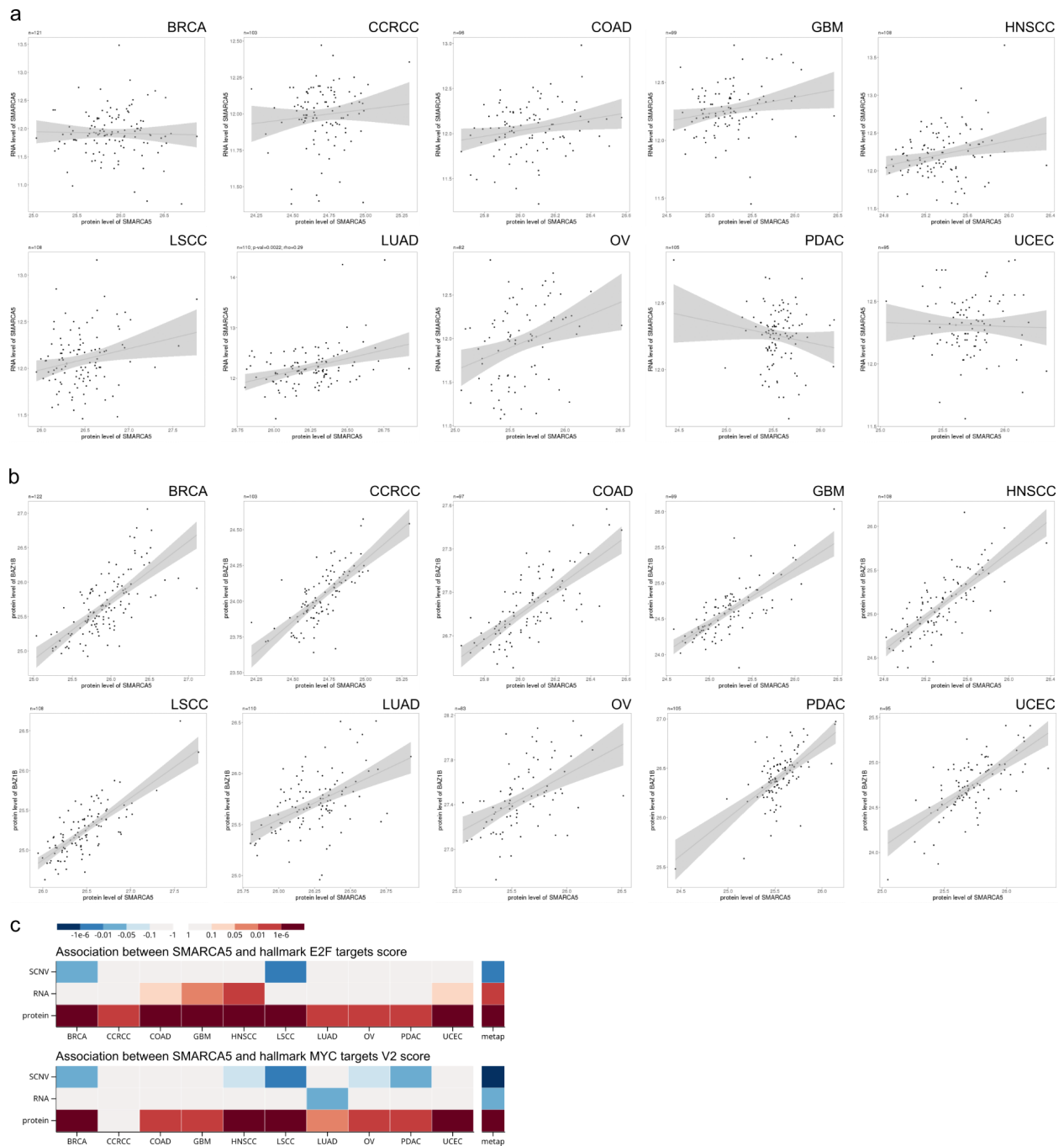


Figure S3. Associations of SMARCA5. (a) Scatter plots of associations between SMARCA5 mRNA abundance and protein abundance in all cancer cohorts. (b) Scatter plots of associations between SMARCA5 protein abundance and BAZ1B protein abundance in all cancer cohorts. (c) Heatmap summarizing p-values of associations between SMARCA5 and E2F targets signature score and MYC targets signature score, respectively. The SMARCA5 protein levels show stronger associations than mRNA.

Protein

Search gene

Gene	Meta P	Signed p-values										
		BRCA	CCRCC	COAD	GBM	HNSCC	LSCC	LUAD	OV	PDAC	UCEC	
+ BMP1	3.6e-9	-	5.6e-5	-	0.41	0.71	0.32	4.3e-4	0.082	1.3e-4	0.17	
+ LOXL2	6e-9	-	3.3e-4	0.35	0.14	0.35	0.081	9.3e-4	0.61	1.5e-3	0.057	
+ NUP35	9.4e-9	-	0.022	0.025	0.073	0.019	4.3e-3	0.11	0.35	0.02	0.32	
+ LMNB2	4.5e-8	-	0.031	0.052	0.037	0.018	0.035	0.03	0.39	0.19	0.17	
+ PLOD1	4.9e-8	-	7.2e-7	0.2	0.062	0.64	0.23	0.11	0.96	6.3e-3	0.03	
+ COLGALT1	1.2e-7	-	3.7e-4	0.46	0.62	0.018	0.089	0.22	0.22	0.014	0.035	
+ PLOD2	1.9e-7	-	1.7e-6	0.11	0.21	0.18	0.16	0.038	0.99	4.5e-3	0.7	
+ GPX2	2.5e-7	-	9.1e-3	0.041	-	0.086	0.23	0.058	-	2.9e-4	0.74	
+ ERBB2	3e-7	-	1.6e-4	0.16	0.042	0.46	0.27	0.36	1.5e-5	0.28	0.99	
+ TRMT1	3.5e-7	-	2.4e-4	0.053	0.12	0.039	3.5e-3	0.39	-	0.74	0.52	

Showing 1 to 10 of 19701 rows

< 1 2 3 4 5 ... 1971 >

Figure S4. Top overall survival analysis results. Top 10 proteins associated with poor prognosis in pan-cancer analysis results.