MOLECULAR BIOLOGY

SUGAR-seq enables simultaneous detection of glycans, epitopes, and the transcriptome in single cells

Conor J. Kearney[1,2]*[†], Stephin J. Vervoort[1,2]*, Kelly M. Ramsbottom[3], Izabela Todorovski[1], Emily J. Lelliott[3], Magnus Zethoven[1], Lizzy Pijpers[1], Ben P. Martin[1], Timothy Semple[1], Luciano Martelotto[4], Joseph A. Trapani[2,3], Ian A. Parish[2,3], Nichollas E. Scott[5], Jane Oliaro[2,3†‡], Ricky W. Johnstone[1,2,6†‡]

Multimodal single-cell RNA sequencing enables the precise mapping of transcriptional and phenotypic features of cellular differentiation states but does not allow for simultaneous integration of critical posttranslational modification data. Here, we describe SUrface-protein Glycan And RNA-seq (SUGAR-seq), a method that enables detection and analysis of N-linked glycosylation, extracellular epitopes, and the transcriptome at the single-cell level. Integrated SUGAR-seq and glycoproteome analysis identified tumor-infiltrating T cells with unique surface glycan properties that report their epigenetic and functional state.

INTRODUCTION

Multimodal single-cell sequencing technologies enable the in-depth characterization of heterogeneous cell populations and have facilitated the identification of unique cell differentiation states within a complex cellular pool (1–4). However, existing technologies do not provide a comprehensive snapshot of all cell phenotypic states, as they fail to incorporate essential posttranslational information that can provide important insights into cell biology (5). Glycosylation, for example, is a carbohydrate-based posttranslational protein modification that regulates multiple cellular processes across diverse cell types and species. Examples include regulation of cell development, differentiation, signal transduction, migration, and immune homeostasis. This occurs through a variety of mechanisms including regulation of protein folding, conformation, distribution, stability, and activity (6). Mammalian protein glycosylation encompasses N-linked glycosylation (N-glycans), O-linked glycosylation (O-glycans), glycosaminoglycans, lipid glycosylation (glycolipids), and glycosyl-phosphatidylinositol-linked proteins. Biosynthesis of N-linked glycoproteins is characterized by the addition of glycans to an asparagine residue and occurs in an ordered, multistepped process that relies on the substrate specificities of glycosyltransferase, which synthesize glycan chains, and glycosidases, which hydrolyze particular glycan linkages. It is the co-operation of both glycosyltransferases and glycosidases that determine the final structural outcome of the glycan branch complexity (7). N-linked glycosylation is critical for T cell signaling and differentiation (8–12), and surface N-glycans may serve as a reporter of distinct states of T cell functionality in cancer, viral infections, and autoimmunity (12–14). Thus, the simultaneous detection of N-glycan levels, surface protein expression, and the transcriptome using existing single-cell platforms would enable deeper phenotypic analysis of diverse cell types across multiple pathophysiology states.

RESULTS

Development of SUrface-protein Glycan And RNA-sequencing

Glycans are derived from complex biological processes and are structurally diverse, which presents a challenge for their detection within the boundaries of current single-cell RNA sequencing (sc-RNA-seq) technology. Lectins, carbohydrate-binding proteins that selectively recognize distinct sugar groups on proteins, were therefore used to report distinct cell surface glycans (15). The lectin *Phaseolus vulgaris agglutinin* (L-Pha) that specifically binds to complex-type N-glycan branches (16, 17) was selected to report glycan abundance on the T cell surface. To first demonstrate the specificity of L-Pha and identify the enzymatic determinants of L-Pha binding to N-glycans on the surface of T cells using an unbiased approach, we performed a genome-wide CRISPR screen in EL4 cells, a T cell line with high levels of baseline N-glycosylation (fig. S1A). EL4 cells negative for complex-type glycans were enriched for guide RNAs targeting multiple genes involved in generating complex-type N-linked glycan branches, including *Mgat1*, *Mgat2*, and *Mgat5* (fig. S1, B and C) (18). The screen results were validated by genetic deletion of selected pathway components (fig. S1D) or chemical inhibition of the alpha-1,3-mannosyl-glycoprotein 2-beta-N-acetylglucosaminyltransferase (Mgat) pathway with swainsonine (fig. S1E). Furthermore, we observed a reciprocal increase in the mannose-binding lectin concanavalin A upon swainsonine treatment or Mgat1 knockout, confirming L-Pha as a specific probe for detecting complex-type N-linked glycan levels on cells (fig. S1, F and G). These data indicated that L-Pha could be used to convert N-glycan abundance on the cell surface into a sequenceable readout at the single-cell level.

SUrface-protein Glycan And RNA-sequencing identifies tumor-infiltrating T cells with unique surface glycan profiles

To develop a modular approach that would be transferrable across multiple platforms, we used biotinylated L-Pha, and secondary detection using an anti-biotin monoclonal antibody conjugated to an

[1]Translational Haematology Program, Peter MacCallum Cancer Centre, Melbourne, Victoria 3000, Australia. [2]Sir Peter MacCallum Department of Oncology, University of Melbourne, Victoria 3010, Australia. [3]Cancer Immunology Program, Peter MacCallum Cancer Centre, Melbourne, Victoria 3000, Australia. [4]Centre for Cancer Research, University of Melbourne, Melbourne 3000, VIC, Australia. [5]Department of Microbiology and Immunology, University of Melbourne at the Peter Doherty Institute for Infection and Immunity, Melbourne 3000, Australia. [6]Cancer Therapeutics Program, Peter MacCallum Cancer Centre, Melbourne, Victoria 3000, Australia.
*These authors contributed equally as first authors.
†Corresponding author. Email: ricky.johnstone@petermac.org (R.W.J.); conor.kearney@petermac.org (C.J.K.); jane.oliaro@petermac.org (J.O.)
‡These authors contributed equally as senior authors.

oligonucleotide was used to integrate glycan status with multimodal sc-RNA-seq technology and enable the simultaneous analysis of global N-glycan levels, protein expression, and the transcriptome. This novel SUrface-protein Glycan And RNA-seq (SUGAR-seq) technique was then applied to examine tumor-infiltrating T cells (TILs), as this cellular pool encompasses diverse T cell differentiation and functional states, including those required for response to immune checkpoint blockade (19). TILs [T cell receptor–positive (TCR⁺), CD3⁺] isolated from B16–ovalbumin (Ova) and MC38-Ova tumors grown in C57BL/6 mice were initially stained with biotinylated L-Pha

(1 μg/ml) combined with nonbiotinylated L-Pha (1:5 ratio) to prevent signal saturation (fig. S2A). Cells were then labeled with antibody-derived tags (ADTs) targeting biotin, the functional markers programmed cell death protein 1 (PD-1) and T-cell immunoglobulin and mucin-domain containing-3 (TIM3), and "hash-tagging" ADTs for demultiplexing TILs from either B16-Ova or MC38 tumors, followed by capture on the 10X Genomics platform (Fig. 1A). TILs from B16-Ova and MC38-Ova tumors were specifically identified by hash tagging (fig. S2B), and uniform manifold approximation and projection (UMAP) clustering using transcriptional signatures identified
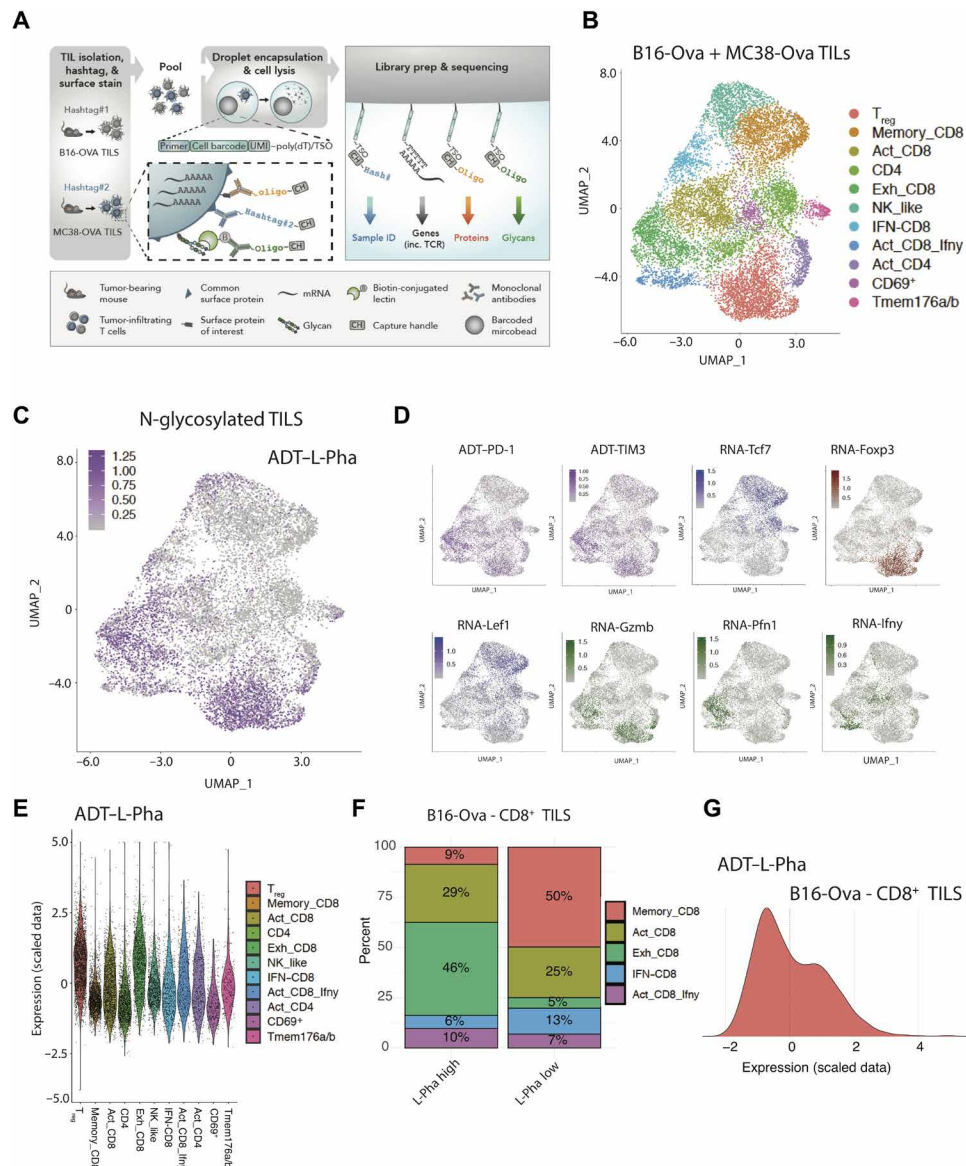

Fig. 1. Development and implementation of SUGAR-seq. (**A**) Schematic representation of SUGAR-seq. TILs (CD3⁺ TCR⁺) were isolated from day 14 B16-Ova and MC38-Ova tumors (*n* = 6, pooled) by FACS. TILs were initially stained with biotinylated (1 μg/ml) and nonbiotinylated L-Pha (1:5 ratio), washed extensively, and then stained with ADT antibodies (anti–PD-1, anti-TIM3, and anti-biotin) and hash-tagging antibodies for sample demultiplexing. Single-cell capture was performed using the 10x Genomics platform. (**B**) UMAP clustering of TILs derived from combined B16-Ova and MC38-Ova tumors based on RNA markers from SUGAR-seq. NK, natural killer. (**C**) UMAP clustering displaying the L-Pha signal (ADT) from SUGAR-seq, on TILs derived from B16-Ova and MC38-Ova tumors combined. (**D**) UMAP clustering from SUGAR-seq displaying the ADT signal for PD-1, TIM3, and RNA expression of particular cluster markers from SUGAR-seq. (**E**) Violin plot of the L-Pha signal (ADT) across the clusters identified in (B). (**F**) TILs were determined to be L-Pha^low or L-Pha^high via separation around a detection (ADT L-Pha) *z* score of 0, followed by CD8⁺ cluster frequency analysis. (**G**) The L-Pha signal (ADT) is displayed surrounding the *z* score on CD8⁺ B16-Ova TILs.

distinct subsets of TILs (Fig. 1B and fig. S2, C and D). Analysis of the L-Pha signal revealed divergent levels of N-glycosylation across distinct TIL populations (Fig. 1C) with highly N-glycosylated TILs enriched in the regulatory T cell (T$_{reg}$) and exhausted T cell subsets, while memory T cells displayed low glycan abundance (Fig. 1, D and E, and fig. S2, E to H). Single CD8$^+$ T cells separated as L-Pha "low" and "high" around the L-Pha signal z score identified enrichment of "exhausted" CD8$^+$ T cells in the L-Pha^{high} compartment, while memory CD8$^+$ T cells were enriched in the L-Pha^{low} compartment (Fig. 1F and fig. S2I). SUGAR-seq was capable of detecting between 2000 and 4000 genes per cell (fig. S3, A to C), and the ADT L-Pha signal represented approximately 80% of total ADT reads compared to PD-1 and TIM3 (fig. S3, D and E). To confirm these observations and to extensively validate SUGAR-seq, we ana-

lyzed N-glycan levels on CD8$^+$ TILs from B16-Ova tumors using a fluorescence-activated cell sorting (FACS)–based assay. We confirmed the presence of a bimodal population of CD8$^+$ TILs for surface N-glycans, as detected by a fluorescein isothiocyanate (FITC)–conjugated form of L-Pha (Fig. 2A). As detected by SUGAR-seq, L-Pha^{high} cells were associated with expression of the exhaustion markers, PD-1 and TIM3, while L-Pha^{low} cells lacked PD-1 and TIM3 expression (Fig. 2B). L-Pha^{low} cells also lacked expression of effector/exhaustion-associated TOX (thymocyte selection-associated high mobility group box protein), TBET (T-box expressed in T cells), and EOMES (eomesodermin) but expressed T cell factor 1 (TCF1), a transcription factor (TF) that marks memory T cells (Fig. 2C). Together, SUGAR-seq accurately identifies distinct cellular subsets with unique N-glycan profiles on a single-cell level.
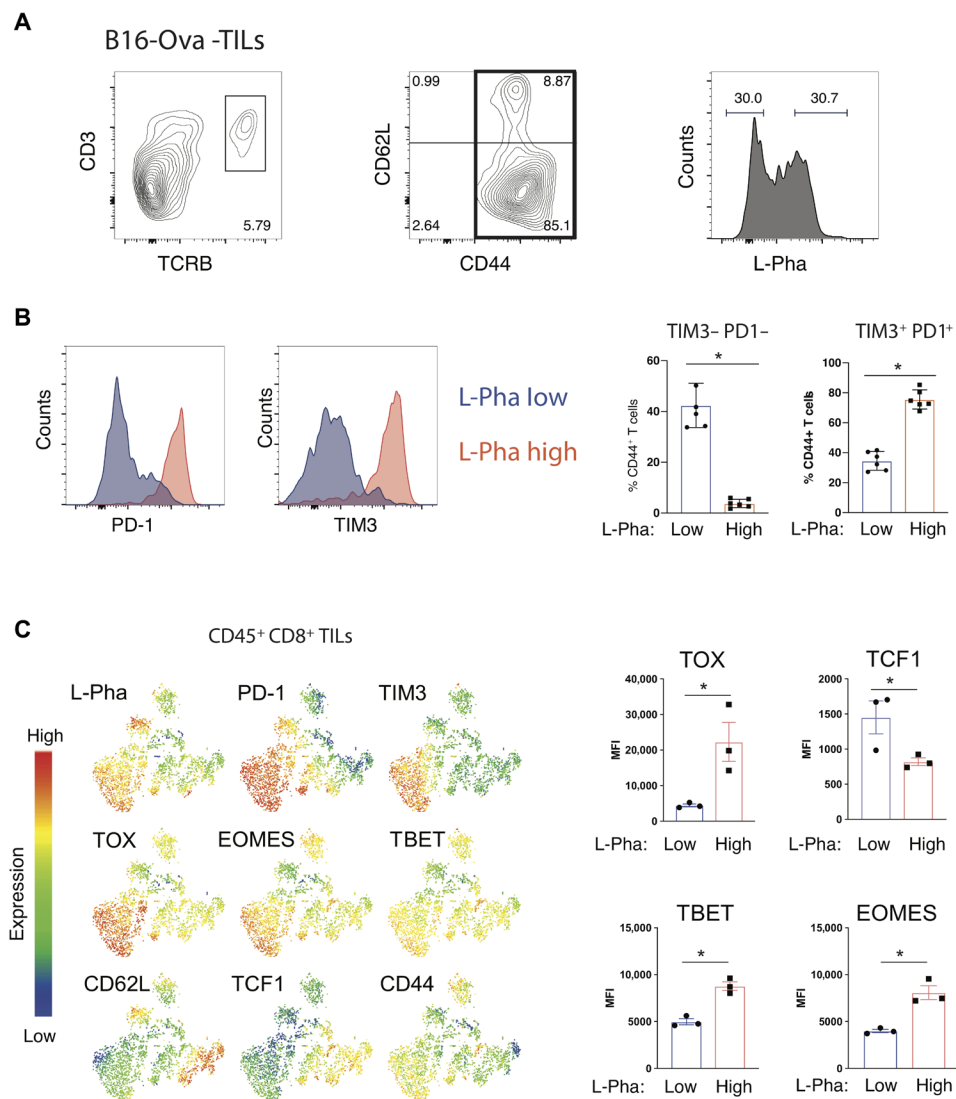

Fig. 2. Validation of SUGAR-seq. (**A**) TILs from day 14 B16-Ova tumors were FACS gated as CD8$^+$ TCR$^+$ CD3$^+$, followed by selecting CD44$^+$ CD62L$^{+/−}$ and then analyzed for N-glycan abundance through detection of FITC-conjugated L-Pha. (**B**) L-Pha^{high} and L-Pha^{low} CD8$^+$ TILs from (A) (right-hand panel) were analyzed for expression of PD-1 and TIM3 by FACS. Bar charts represent the frequency of PD-1 TIM3–positive or PD-1 TIM3–negative CD8$^+$ TILs across L-Pha^{low} or L-Pha^{high} subsets. Data points represent individual mice. *$P < 0.01$. (**C**) TILs (CD45$^+$ CD8$^+$) from day 14 B16-Ova tumors were FACS analyzed for the indicated surface proteins, L-Pha staining, and intracellular proteins as indicated. TSNE clustering is depicted. Bar charts represent the mean fluorescence intensity (MFI) of the indicated TFs in CD45$^+$ CD8$^+$ TILs across L-Pha^{low} or L-Pha^{high} subsets. Data points represent individual mice. *$P < 0.01$.

SUGAR-seq enables detection of key T cell regulatory molecules

To identify which T cell surface proteins bound L-Pha, we performed lectin-based proximity labeling (20) followed by proteomic analysis using wild-type EL4 cells or *Mgat1* CRISPR knockout cells as a control. This identified that multiple T cell surface molecules were labeled by L-Pha in an *Mgat1*-dependent manner (figs. S4, A to D). Consistent with the surface specificity of L-Pha labeling, gene ontology (GO) term analysis revealed an enrichment of proteins involved in cellular adhesion and leukocyte activation. This included key glycoprotein regulators of T cell activity, including Pdcd1 (programmed cell death protein 1), Tcrb, and Cd28 (fig. S4, C to E). To assess the impact of *Mgat1* on the glycoproteome and proteome levels, we performed label-free quantitative (LFQ) analysis of control and *Mgat1* CRISPR knockout EL4 cells. Few alterations within the proteome were noted in *Mgat1* knockout cells, yet at the glycoproteome level, marked changes in N-linked glycan compositions were observed. Analysis of the N-linked glycans observed on glycopeptide enriched revealed a marked increase in HexNAc(2)Hex (5) within *Mgat1* knockouts compared to wild-type EL4 cells (fig. S5, A and B). To further confirm these changes independent of L-Pha–based approaches, amino-oxy-biotin surface labeling was undertaken followed by LFQ proteomic analysis. This revealed a reduction in labeling of several important T cell surface molecules including Pd1, Icam1, Cd28, and poliovirus receptor (Pvr) (fig. S6A). GO term analysis revealed a significant enrichment in proteins associated with adhesion, migration, and T cell activation (fig. S6B). These alterations in composition are consistent with the glycan alterations observed from chemical inhibition of complex-type N-linked glycan formation with swainsonine as monitored by tandem mass tag (TMT)–based quantitative proteomes and glycoproteome (fig. S7). At the individual glycoprotein level, this alteration in glycans from complex to high-mannose N-glycans was observed on key T cell regulatory molecules, including Cd8a, Gzma, and cathepsin Z (fig. S7). Combined, these experiments demonstrate that surface L-Pha binding enables the measurement of complex-type N-linked glycan status within T cells.

SUGAR-seq in the context of health and disease

SUGAR-seq was used to compare the complex-type N-glycan profile of T cells from the lymph node of a healthy mouse, T cells from the lymph node of a mouse bearing a subcutaneous B16 tumor, and TILs from this tumor-bearing mouse (Fig. 3A). As expected, we found divergent CD4+ and CD8+ cluster frequencies in lymph nodes and in the tumor (Fig. 3, B and C). *Foxp3*-positive T$_{regs}$ could be found in both lymph nodes in the tumor; *Sell* and *Nkg7*+ cells were largely found in the lymph nodes, while Ifng+ cells were confined to the tumor (Fig. 3D). We found that the ADT L-Pha signal was significantly increased in the TILs compared to those from lymph nodes (Fig. 3E), suggesting that cellular microenvironment modulates complex-type N-glycan levels. Across all clusters, T$_{regs}$ and activated/exhausted CD8+ T cells displayed the most extensive L-Pha binding, while naïve CD4+ cells and memory CD4/CD8+ cells displayed little L-Pha binding (Fig. 3F).

SUGAR-seq identifies distinct cellular states of differentiation

To further characterize the identity of cells exhibiting divergent glycan levels, L-Pha^{high} and L-Pha^{low} CD8+ TILs from B16-Ova tumors were isolated and analyzed by bulk RNA-seq and assay for transposase-accessible chromatin using sequencing (ATAC-seq; Fig. 4A). ATAC-seq revealed that L-Pha^{high} TILs displayed increased global chromatin accessibility, particularly at loci (e.g., *Gzmb*) associated with effector function/exhaustion, while L-Pha^{low} TILs had increased accessibility at loci associated with a T cell memory phenotype (Fig. 4B and fig. S8A). Consistent with our RNA-seq data, TF motif analysis revealed enrichment of the exhaustion-associated TFs, BATF and activating protein 1, in L-Pha^{high} TILs, while the memory-associated TFs, TCF and FLI-1 (friend leukemia integration 1 transcription factor -1), were enriched in L-Pha^{low} TILs (fig. S8B). Bulk RNA-seq revealed concordance between gene expression and open chromatin in both L-Pha^{high} (*Havcr2*, *Serpine2*, *Gzme*, *Gzmf*, and *Prdm1*) and L-Pha^{low} (*Tcf7*, *Cxcr3*, *Cd7*, *S1pr1*, and *Lef1*) TILs (Fig. 4, C to E, and fig. S8, C and D). Gene set enrichment analysis (GSEA) analysis revealed effector/exhausted signatures in L-Pha^{high} cells and memory signatures in L-Pha^{low} cells (Fig. 4F), which was also evident from the RNA-seq and ATAC-seq overlap signature (fig. S8D). Functional analyses of L-Pha^{high} and L-Pha^{low} TILs demonstrated that L-Pha^{high} cells produced higher levels of interferon-γ (IFN-γ) upon restimulation than L-Pha^{low} cells; however, L-Pha^{high} high cells were less efficient at coproducing tumor necrosis factor (TNF)/interleukin-2 (IL-2), TNF/IFN-γ, and IFN-γ/IL-2, consistent with their exhausted phenotype (fig. S9, A to C). Similarly, L-Pha^{high} cells expressed elevated granzyme B and Ki67, indicative of their proliferative nature (fig. S9D).

Pseudobulk gene signatures generated from SUGAR-seq data revealed considerable overlap with the bulk RNA-seq, confirming the specificity of the SUGAR-seq glycan signal (Fig. 4, G and H). Moreover, pseudotime analysis predicted that CD8+ T cells transitioned from the memory T cell cluster to activated and lastly exhausted states of differentiation (Fig. 5A). Overlaying bulk RNA-seq L-Pha signatures with SUGAR-seq CD8+ T cell clusters revealed enrichment of the L-Pha^{high} signal in exhausted T cells over L-Pha^{low} memory T cells (Fig. 5B). This finding was confirmed using published effector and memory CD8+ T cell signatures (Fig. 5C and fig. S10A). Furthermore, the L-Pha signal increased over pseudotime, concordant with PD-1 and TIM3 expression, suggesting that glycan levels are dynamic and altered according to the T cell differentiation state (Fig. 5D and fig. S10B).

Integration of SUGAR-seq and TCR clonality

Integrated TCR-seq and SUGAR-seq data revealed that highly abundant clones were enriched in the activated and exhausted CD8+ T cells, with few clonal cells found in the memory T cell population in both B16-Ova and MC38-Ova tumors (Fig. 5E and fig. S10, C and D). L-Pha^{high} CD8+ T cells were associated with increased clonal expansion, suggesting that tumor reactivity results in elevated complex-type N-glycans that are detected by L-Pha on the T cell surface (Fig. 5E and fig. S10, D and E). These findings were confirmed using Ova-specific tetramer to detect tumor-reactive T cell clones in B16-Ova tumors by FACS, which demonstrated that tetramer-positive TILs were generally L-Pha^{high} and expressed PD-1 and TIM3 (fig. S11, A to C). Consistent with the TCR-seq data, tetramer-positive TILs had higher levels of L-Pha than tetramer-negative TILs and also expressed higher levels of PD-1 and TIM3 (fig. S11, D and E), confirming that tumor reactivity results in an altered glycan state. Last, we analyzed the TCR clonality of the CD4+ clusters (Fig. 5, F and G) and found that clonal CD4+ cells existed predominantly across the T$_{reg}$, CD4+, and Cd69+ clusters (Fig. 5H). We found that clonal T$_{reg}$ and activated CD4+ T cells were higher for the L-Pha signal, mirroring our findings with CD8+ T cells (Fig. 5H).
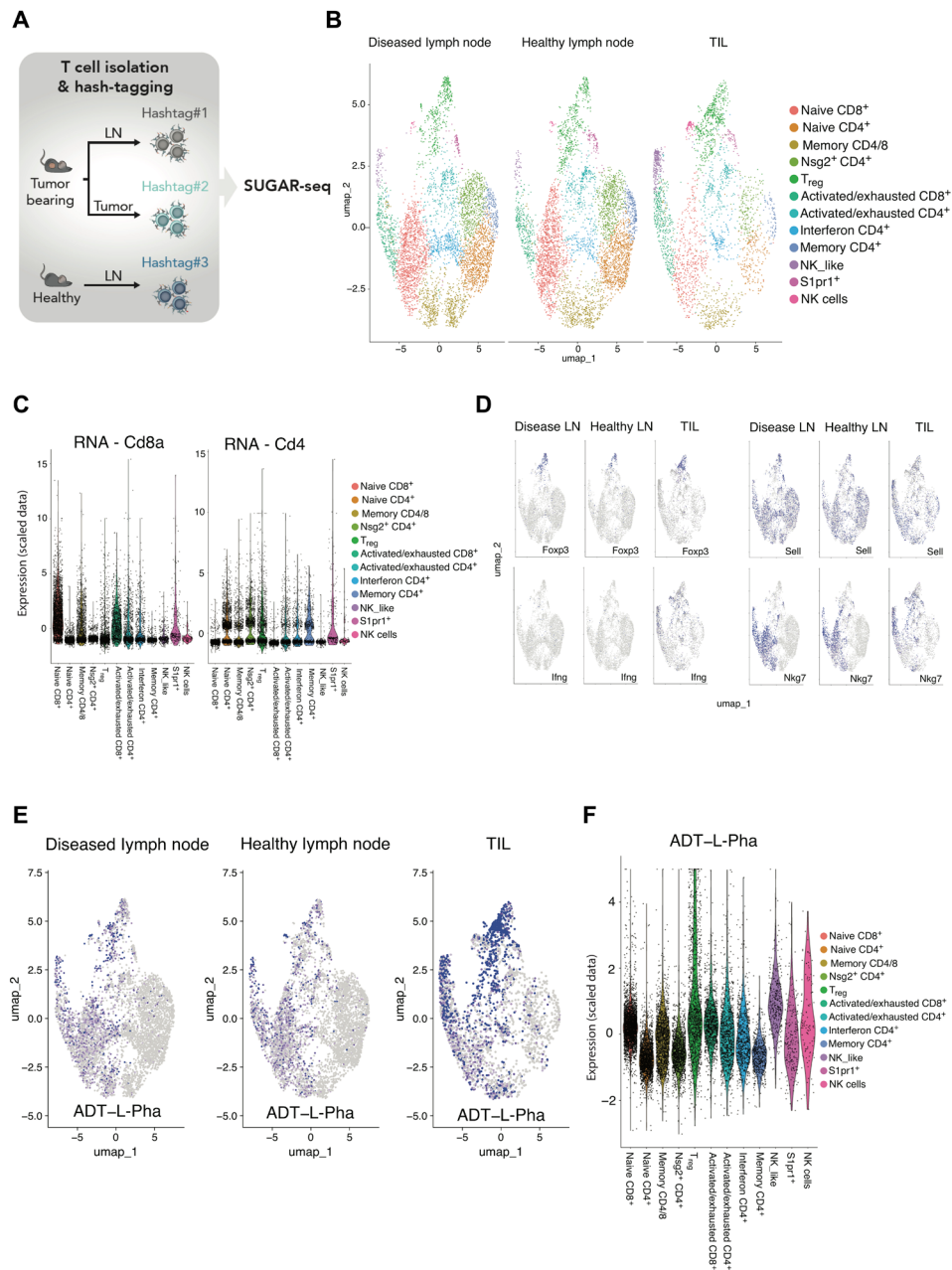
Fig. 3. SUGAR-seq analysis of lymph node–derived T cells. (**A**) Schematic representation of SUGAR-seq experimental design. T cells (CD3$^+$ TCR$^+$) were isolated from the lymph node (LN) of a healthy C57BL/6 mouse, the lymph node of a C57BL/6 mouse carrying a day 14 B16 subcutaneous tumor (diseased lymph node), or the TILs from that mouse by FACS. Cells were initially stained with biotinylated (1 μg/ml) and nonbiotinylated L-Pha (1:5 ratio), washed extensively, then stained with anti-biotin ADT and hash-tagging antibodies for sample demultiplexing. Single-cell capture was performed using the 10x Genomics platform. (**B**) UMAP clustering of populations derived from combined samples described in (A) based on RNA markers. (**C**) Violin plots of CD4 and CD8a transcript levels across the UMAP clusters described in (B). (**D**) UMAP plots displaying the expression of the indicated genes across diseased lymph node, healthy lymph node, and TIL conditions. (**E**) UMAP plots displaying the ADT–L-Pha signal genes across diseased lymph node, healthy lymph node, and TIL conditions. (**F**) Violin plot of ADT–L-Pha signal across the UMAP clusters described in (B).

DISCUSSION

Here, the application of SUGAR-seq has enabled the simultaneous detection of surface glycans, epitopes, transcripts, and TCR repertoire to identify TILs with distinct differentiation states and functional capacities. Beyond this, SUGAR-seq has future wide-ranging application if coupled to perturbations such as RNA interference, single-cell CRISPR screening, or other genetic or pharmacological perturbation methods, which would rapidly improve our understanding of

glycobiology across diverse biological and experimental systems. SUGAR-seq can be implemented to identify cellular phenotypes associated with diseases including cancer, autoimmunity, and obesity. Given that biotinylated lectins are widely available from commercial sources, SUGAR-seq can be readily performed with alternative lectins that detect distinct forms of both O-linked and N-linked glycosylations or alternative posttranslational modifications. Note, however, that the ability of particular lectins such as L-Pha may be
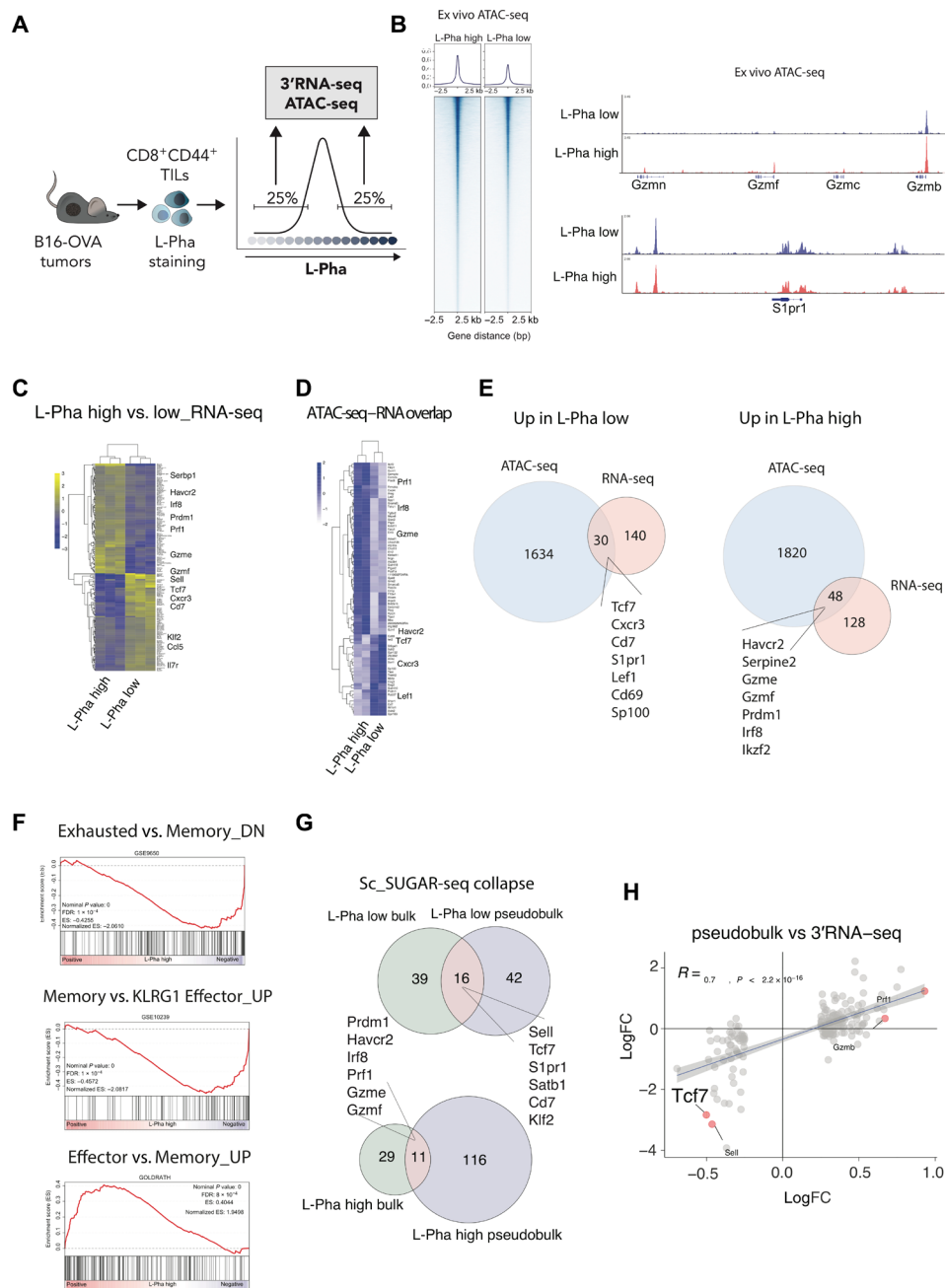
Fig. 4. SUGAR-seq identifies TILs with divergent epigenetic and functional capacities. (**A**) Schematic representation of ex vivo RNA-seq and ATAC-seq. TILs (CD3[+] TCR[+] CD8[+] CD44[+]) were FACS sorted on L-Pha[low] and L-Pha[high] from day 14 B16-Ova tumors (three pools of six mice), followed by 3′ bulk RNA-seq and ATAC-seq analysis. (**B**) Chromatin accessibility heatmap for L-Pha[low] and L-Pha[high] TILs from (A). Chromatin accessibility peaks for the granzyme cluster locus and the S1pr1 locus are displayed on the right. (**C**) RNA-seq heatmap for significantly differentially regulated genes comparing L-Pha[low] and L-Pha[high] TILs from (A). (**D**) Heatmap of overlapping significantly differentially regulated genes from ATAC-seq and RNA-seq comparing L-Pha[low] and L-Pha[high] TILs from (A). (**E**) Venn diagram analysis of overlapping significantly differentially regulated genes from ATAC-seq and RNA-seq comparing L-Pha[low] and L-Pha[high] TILs from (A). (**F**) GSEA analysis of the RNA-seq analysis from (A). (**G**) Overlapping significantly differentially regulated genes from the bulk RNA-seq (A) and pseudobulk analysis significantly differentially regulated genes from SUGAR-seq data (Fig. 1). (**H**) Correlation between aggregated single-cell sequencing (pseudobulk) and ex vivo bulk RNA-seq. FC, fold change.

hindered in their ability to bind complex-type N-glycans if extensive sulfation or sialylation is present. Nevertheless, L-Pha as a single modality identifies TILs with distinct phenotypic and functional capacities. Given the abundance of glycans on the surface of cells, lectins may be an economic and convenient method to hashtag samples within a single capture reaction. Furthermore, as glycosylation is conserved in multi-

cellular organisms, SUGAR-seq can be used to study glycan biology in diverse species without the need for specialized reagents, as is the case with antibody-mediated detection of epitopes. SUGAR-seq is a new and adaptable technology that provides true multiomic data encompassing glycan status with sc-RNA-seq and is readily adaptable to other microfluidic or microwell platforms, making this approach widely accessible.
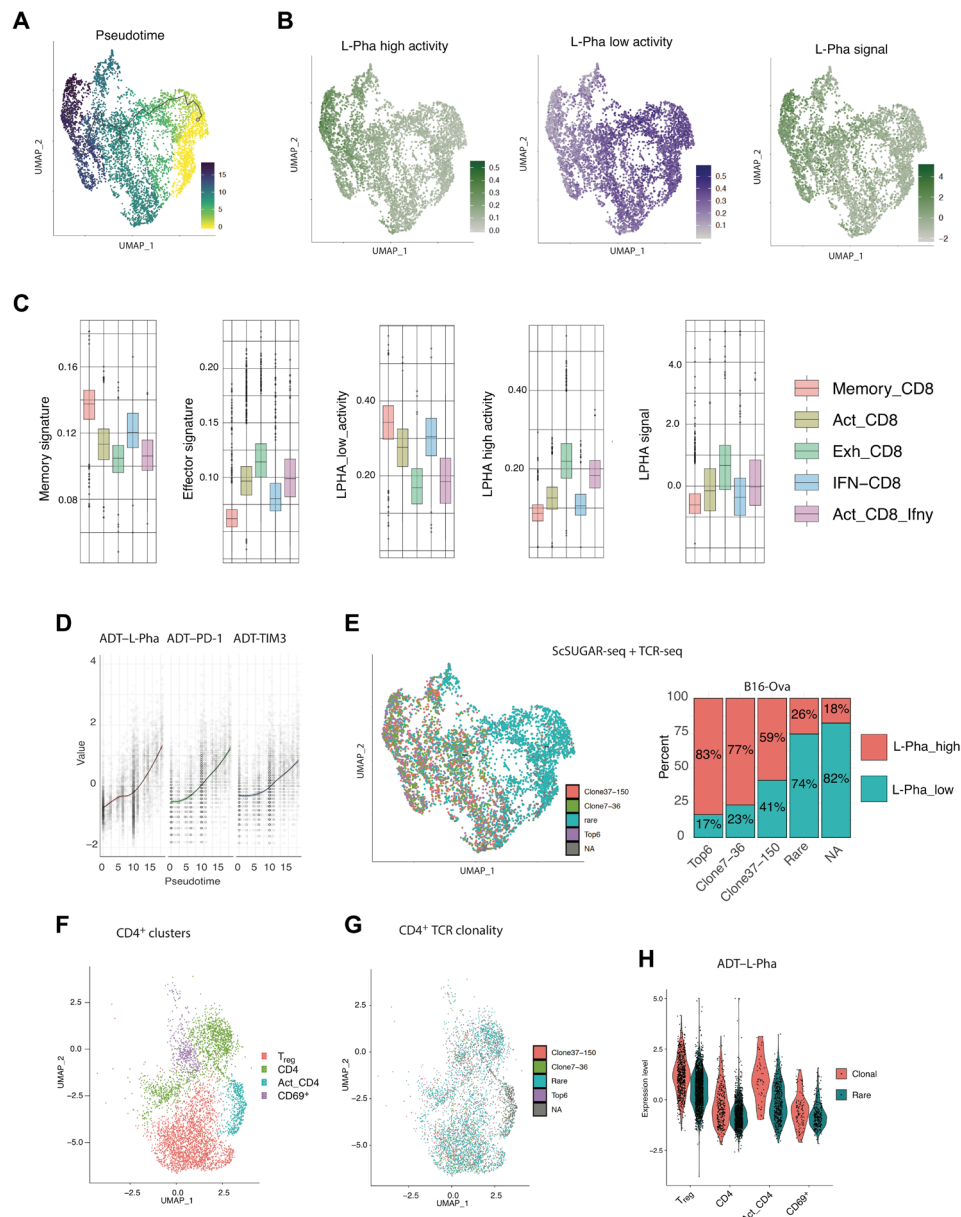
Fig. 5. Integration of SUGAR-seq and TCR-seq reveals that tumor reactivity alters the N-glycosylation state of T cells. (**A**) Monocle was used to calculate pseudotime analysis of CD8[+] TIL clusters from pooled B16-Ova and MC38-Ova tumors from SUGAR-seq. (**B**) L-Pha[low] and L-Pha[high] gene signatures were generated from the bulk RNA-seq (A) and visualized on SUGAR-seq CD8[+] T cell clusters using AUCell. The L-Pha ADT signal is visualized in the right-hand panel. (**C**) Boxplots of AUCell analysis of memory and effector gene signatures (GOLDRATH_EFF_VS_MEMORY_CD8_TCELL) and bulk RNA-seq–derived L-Pha signatures (B) across CD8[+] T cell clusters. (**D**) Monocle pseudotime analysis of ADTs L-Pha, PD-1, and TIM3 on CD8[+] TIL clusters from pooled B16-Ova and MC38-Ova tumors (SUGAR-seq). (**E**) TCR-seq clonotypes for L-Pha[low] and L-Pha[high] CD8[+] TILs (separated around an ADT L-Pha z score of 0) from pooled B16-Ova and MC38-Ova tumors. Right-hand panel: Clonotype frequency of CD8[+] TILs from L-Pha[low] and L-Pha[high] subpopulations, derived from B16-Ova tumors. (**F**) UMAP of CD4 transcript–expressing clusters from B16-Ova TILs. (**G**) TCR-seq clonotypes across CD4[+] UMAP clusters described in (F). (**H**) Violin plot of the ADT–L-Pha signal across Cd4[+] UMAP clusters, split by "clonal" (Top6 + Clone7–36 + Clone37–150) and "rare" (rare + NA).

MATERIALS AND METHODS

In vivo experiments

Animal work was conducted with approval from the Peter MacCallum Animal Experimentation Ethics Committee. C57BL/6 mice were purchased from Walter and Eliza Hall Institute (WEHI). B16 and MC38 cells were engineered to express Ova, and tumors were injected subcutaneously on the right flank. On day 14, tumors were digested with collagenase IV (1.6 mg/ml) and deoxyribonuclease

(2 U/ml) in Dulbecco's modified Eagle's medium (DMEM) for 45 min at 37°C with agitation and filtered through a 70 μM filter for staining with FACS antibodies.

Flow cytometry and cell sorting

Mouse T cells were isolated from tumors for SUGAR-seq by FACS on CD3[+]TCRB[+]CD90.2[+]CD11b[−] cells. For ex vivo RNA-seq and ATAC-seq, cells were sorted on CD3[+]TCRB[+]CD44[+]CD8[+]. Fixable

yellow (Invitrogen, L34959) or propidium iodide was used to stain live/dead cells. Anti-mouse antibodies used were as follows: CD3 (17A2), TCRB (H57-597), CD11b (M1/70), CD90.2 (53-2.1), CD8α (clone 53-6.7), CD44 (clone IM7), CD62L (clone MEL14), CD45.2 (clone 104), TCF1 (C63D9), PD-1 (29F.1A12), TIM3 (B8.2C12), Ki67 (B56), EOMES (W17001), TBET (4B10), GZMB (GRB04), IFN-γ (XMG1.2), TNF-α (MP6-XT22), and IL-2 (JES6-5H4). Cell sorting was conducted using BD FACSAria Fusion and analysis was performed on BD LSRFortessa X-20 or BD FACSymphony flow cytometer (BD Biosciences, North Ryde, New South Wales, Australia). Data were analyzed using FlowJo LLC software.

Cell lines
MC38-OVA and B16-OVA cell lines were cultured in DMEM containing 10% fetal bovine serum (FBS) and cultured at 37°C in 10% CO_2. EL4 cells were cultured in DMEM containing 10% FBS and 1% GlutaMAX.

Lectins
Biotinylated L-Pha was purchased from Vector Labs (B-1115). FITC-conjugated L-Pha was purchased from Invitrogen (L11270).

Primary mouse T cell isolation and culture
Naïve CD8$^+$ T cells were isolated from C57BL/6 mouse spleens using EasySep Mouse CD8$^+$ T Cell Isolation Kit (STEMCELL Technologies) and labeled with division tracking dye, CellTrace Violet. Purity of T cell population was verified as >95% CD8$^+$ CD4$^-$ by flow cytometry. Labeled CD8$^+$ T cells were stimulated with plate-bound αCD3 antibody, αCD28 antibody (clone 37.51) (2 μg/ml; WEHI), and mIL-2 (100 U/ml; WEHI). All primary mouse cells were cultured in RPMI 1640 with 20 mM Hepes containing 10% FBS, 1% GlutaMAX, 1 mM sodium pyruvate, 1 mM MEM non-essential amino acids, 0.1% 2-mercaptoethanol, and antibiotic-antimitotic and cultured at 37°C in 5% CO_2 and supplemented with IL-2 (100 IU/ml).

SUrface-protein Glycan And RNA-seq
FACS-sorted TILs were incubated with a mixture of biotinylated and nonbiotinylated L-Pha (1:5 ratio) at a concentration of 1 μg/ml for 20 min on ice. After three washes with cold phosphate-buffered saline (PBS), cells were stained with cell hashing antibodies and CITE-seq (cellular indexing of transcriptomes and epitopes by sequencing) antibodies as previously described (1). "Stained" and washed cells were counted, brought to ~1000 cells/μl, and loaded onto the 10× chromium instrument (10x Genomics, Pleasanton, CA, USA) to generate single-cell gel beads in emulsion and capture/barcode cells. For 5′ multiomics including gene expression, TCR-sequencing and hashing/CITE-seq, we used the 10× Single Cell V(D)J Kit with Feature Barcoding (enabled for 5′ gene expression, TCR, and feature barcoding for cell surface protein) protocol following the manufacturer's instructions. TCR libraries were prepared using the 10x Chromium Single Cell V(D)J Enrichment Kit, mouse T cell. 5′ hashtag oligos (HTO)/ADT (one library) was prepared using the Chromium Single Cell 5′ Feature Barcode Library Kit and indexed using the Chromium i7 Multiplex Kit N, Set A. Cell Ranger reads were aligned to the mm10 reference genome; cellular barcodes were demultiplexed, and unique molecular identifiers and antibody capture (ADT and HTO) were quantified using 10x Genomics' Cell Ranger software (version 3.1.0). Cell barcodes containing RNA or antibody counts from cells from more than one sample (intersample doublets) were identified using Seurat's HTODemux function. Barcodes containing counts from more than one cell within the same sample (intrasample doublets) were identified using the Scrublet Python package (version 0.2.1) (21). A cutoff of more than one median absolute deviation value above the median Scrublet score was chosen. Cells identified as either type of doublet were removed from the analysis. Relative TF activity in each cell was estimated using the single-cell regulatory network inference and clustering method (22). Area-under-curve (AUC) scores for TFs were calculated on the basis of inferred gene regulatory networks using the pySCENIC Python package (version 0.9.19). Gene expression and antibody count matrices were processed in R (version 3.6.1) using the Seurat R package (version 3.1.0) (23). RNA transcript counts for barcodes identified as cells by Cell Ranger were normalized using sctransform via Seurat's SCTransform function. The counts were first transformed with no covariates in the sctransform model; then, cell cycle phase scores were estimated using Seurat's CellCycleScoring function with mouse homologs of the cell cycle gene sets provided by Seurat. The sctransform normalization was then rerun with the cell cycle phase scores and the percentage of raw RNA counts belonging to mitochondrial genes for each cell as variables to be regressed out in the model. ADT counts were log1p transformed and scaled (z score) or normalized using centered log ratio transformation with either method, producing a similar result. Principal components analysis was then performed on the sctransform-scaled RNA expression values for genes with residual variance in the sctransform model greater than 1.3. A shared nearest neighbor (SNN) network was calculated using the top 10 principal components using the FindNeighbors function with k-nearest neighbors set to 50 and cosine distance metric. The SNN network was then used to identify cell populations using the FindClusters function using the Louvain algorithm with resolution parameter 0.6. UMAP values were also calculated using the RunUMAP function with the top 10 principal components as input and parameters n.neighbors = 50 and metric = "cosine." Monocle3 was used for pseudotime analysis using default setting.

Whole-genome CRISPR screen
Genome-wide CRISPR screens were performed as described previously (24). EL4 cells were transduced with mCherry-Cas9 using the (FUCas9Cherry) and cell sorted for Cherry$^+$ cells. mCherry-Cas9–expressing cells were then transduced with lentivirus containing the Brie genome-wide single guide RNA (sgRNA) library. Forty-eight hours after transduction, successfully transduced cells were selected with puromycin (1 mg/ml; Millipore) for 5 days. After selection at the commencement of screen, a reference sample at time point zero, T0, was snap-frozen and stored at −80°C. For the screen, library-containing cells (20×10^6) were stained with L-Pha–FITC and FACS sorted on the lowest 10th percentile of L-Pha signal. Cells were put back in culture for 5 days; then, this process repeated two more times. Genomic DNA was extracted from screen and control samples using the DNeasy Blood & Tissue Kit (QIAGEN) according to the manufacturer's instructions. sgRNA sequences were then amplified for next-generation sequencing by polymerase chain reaction (PCR) using specific adaptor sequences. PCR products were pooled and cleaned up using the AMPure XP-PCR purification system (Beckman Coulter) according to the manufacturer's instructions. Samples were subsequently multiplexed and sequenced on the NextSeq 500 (Illumina) in-house at the Peter MacCallum Molecular Genomics Core Facility,

generating 75–base pair (bp) single-end reads. After demultiplexing with CASAVA (v1.8), adaptor sequences were removed using Cutadapt (v1.7), leaving the 20-bp sgRNA sequence. Model-based Analysis of Genome-wide CRISPR-Cas9 Knockout v0.5.9 (MAGeCK count and MAGeCK test) software was subsequently used to count the reads and perform sgRNA enrichment and statistical analyses between treated/sorted and control samples. Screen data were visualized using the R package ggplot2 v2.2.1.

CRISPR-CAS9 KNOCKOUT
Targeted gene deletion was performed by electroporation of Cas9 nuclease/synthetic guide RNA (Synthego) into target cells using Amaxa 4D-Nucleofector electroporation system (Lonza, no. AAF-1002B) as per the manufacturer's protocol.

Western blot
Cell pellets were lysed in 2% SDS buffer (0.5 mM EDTA and 20 mM Hepes), boiled (95°C, 5 min), and quantified using the DC protein assay (Bio-Rad) as per the manufacturer's protocol. Equal amounts of total protein in 5× SDS sample buffer [313 mM tris-HCl (pH 6.8), 50% (v/v) glycerol, 10% (v/v) β-mercaptoethanol, 10% (w/v) SDS, and 0.05% (w/v) bromophenol blue] were boiled (95°C, 5 min), loaded into precast gels (Bio-Rad), and resolved via SDS–polyacrylamide gel electrophoresis with running buffer containing 25 mM tris, 190 mM glycine, and 0.1% (w/v) SDS. Precision Plus Protein dual-color standard (Bio-Rad) was used as a molecular weight marker. Proteins were transferred onto methanol-activated polyvinylidene difluoride membranes (Millipore) using the Trans-Blot Turbo semidry transfer system (Bio-Rad) for 20 to 30 min with tris-glycine transfer buffer [50 mM tris, 40 mM glycine, 0.375% (w/v) SDS, and 20% (v/v) methanol]. Membranes were blocked in 5% skim milk made up in tris-buffered saline containing 0.1% Tween 20 (TBS-T) for 1 hour before probing with primary antibody overnight at 4°C, followed by corresponding horseradish peroxidase (HRP)–conjugated secondary antibody for 1 hour at room temperature. Immunoblots were washed thrice with TBS-T (10 min each) after each antibody incubation, and proteins were detected using ECL Western blotting substrate (Amersham GE Healthcare).

3′ RNA-seq
Cells were collected and washed once with ice-cold PBS before resuspension in TRIzol (Thermo Fisher Scientific, 15596026). RNA was isolated using the Direct-zol RNA MiniPrep Kit (Zymo Research, R2052) according to the manufacturer's instructions. Sequencing libraries were prepared using the QuantSeq 3′-mRNA Seq Library Prep Kit for Illumina (Lexogen, Vienna, Austria). Libraries were sequenced on the Illumina NextSeq 500 to obtain 75-bp single-end reads. Sequencing files were demultiplexed using Bcl2fastq (v2.17.1.14) to generate FASTQ files on which quality control (QC) was performed using FASTQC (v0.11.5). Sequencing reads were trimmed using Cutadapt (v1.7) and aligned to the human reference genome (Hg19) using HISAT2 (v2.1.0). Read counting across genomic features was performed using featureCounts (from the Subread package v1.5.0) before differential gene expression analysis using Voom/Limma. GSEA were performed using GSEA software (Broad Institute), and barcode plots were plotted using ReplotGSEA.R (part of the Rtoolbox R package).

Assay for transposase-accessible chromatin using sequencing
Cells were washed once in ice-cold PBS and lysed in ATAC lysis buffer [0.1% Tween 20, 0.1% NP-40, 3 mM MgCl$_2$, 10 mM NaCl, and 10 mM tris-HCl (pH 7.4)]. Tagmentation was performed with Tn5 transposase and 2× TD Buffer (Nextera DNA Library Prep Kit, Illumina) for 30 min at 37°C. Tagmented DNA was purified using a MinElute column (QIAGEN, 28004) and amplified for 12 cycles using 2× KAPA HiFi HotStart ReadyMix (Kapa Biosystems, KK2602). The amplified libraries were purified using MinElute columns (QIAGEN) and sequenced on an Illumina NextSeq 500 with 75-bp single-end reads. Library QC and quantification were performed using D1000 high-sensitivity screen tape with the 4200 TapeStation Instrument (Agilent Technologies) and size selected for between 200 and 700 bp using a Pippin Prep system (Sage Science).

ATAC-seq analysis
Bcl2fastq v2.17.1.14 was used for demultiplexing. The FASTQ files generated by sequencing were aligned to the mouse reference genome (GRCm38/mm10) using bowtie (v2.2.3). Samtools (v1.8) was used for manipulation of SAM and BAM files, after which MACS (v2.0.10) was used for peak calling. Browser-viewable TDF files were generated using IGVTools (v2.3.72), and chromatin immunoprecipitation sequencing tracks were visualized using IGV (v2.3.55). Differentially accessible regions were quantitative analyzed using Rsubread featureCounts on merged reference bed file containing all peaks identified across treatment conditions, after which the Limma-Voom package was used for statistical analysis of differentially accessible regions. Subsequently, HOMER (v4.8.3) was used for motif analysis on MACS2 peak summits using findMotifGenome.pl for differentially accessible regions using all summits as a background set with the -bg option. ATAC-seq peaks were annotated to genes using the annotatePeaks.pl function, after which R was used for visualization.

Proteomics preparation of whole-cell samples
Snap-frozen cells were lysed in guanidine hydrochloride (GdHCl) lysis buffer [6 M GdHCl, 100 mM tris (pH 8.0), 10 mM tris(2-carboxyethyl) phosphine, and 40 mM 2-chloroacetamide] at 95°C and 2000 rpm according to Humphrey et al. (25). Lysates were cooled, and protein concentrations were determined using a bicinchoninic acid assay (Thermo Fisher Scientific, Waltham, MA, USA). One milligram of each protein sample was acetone-precipitated (8 volumes of acetone, 1 volume of water, and 1 volume of sample) overnight at −20°C. Precipitated material was then pelleted at 16,000g for 10 min at 4°C; the supernatant was removed, and the pellets were dried at 65°C to remove residual acetone. Protein pellets were allowed to cool to room temperature and then resuspended in 6 M urea, 2 M thiourea with 10 mM dithiothreitol (DTT), and 100 mM ammonium bicarbonate (ABC). The samples were then incubated at room temperature in the dark for 1 hour to reduce disulfide bonds. Reduced samples were then alkylated by the addition of 50 mM chloroacetamide and incubated for another hour in the dark. Alkylation was halted by the addition of 50 mM DTT and incubated at room temperature for 15 min. Samples were then digested with Lys-C [1:200 (w/w); Wako Lab Chemicals, Japan] for 4 hours at room temperature before dilution with 100 mM ABC and digestion with trypsin [1:50 (w/w); Promega, Madison, WI, USA] overnight at 25°C. Digested samples were acidified to a final concentration of 0.5% formic acid and

desalted with 50 mg of tC18 Sep-Pak columns (Waters Corporation, Milford, MA, USA) according to the manufacturer's instructions. tC18 Sep-Pak columns were conditioned with 10 bed volumes of buffer B (0.1% formic acid and 80% acetonitrile) and then equilibrated with 10 bed volumes of buffer A* [0.1% trifluoroacetic acid (TFA) and 2% acetonitrile] before use. Samples were loaded onto equilibrated columns; then, columns were washed with at least 10 bed volumes of buffer A* before bound peptides were eluted with buffer B. For TMT experiments, eluted peptides were aliquoted into 100-μg aliquots. For label-free glycopeptide/proteome analysis, eluted peptides were aliquoted into a 10-μg aliquot for proteome analysis, and the remaining 990 μg was aliquoted for ZIC-HILIC (zwitterionic hydrophilic interaction liquid chromatography) glycopeptide enrichment. All aliquots were then dried down by vacuum centrifugation.

ZIC-HILIC glycopeptide enrichment

ZIC-HILIC glycopeptide enrichment was performed according to the protocol of Mysling *et al.* (*26*). Briefly, ZIC-HILIC stage tips (*27*) were created by packing 0.5 cm of 10-μm ZIC-HILIC resin (Millipore, MA, USA) into p200 tips containing a frit of C8 Empore (Sigma-Aldrich) material. Before use, the columns were washed with ultrapure water, followed by 95% acetonitrile, and then equilibrated with 80% acetonitrile and 1% TFA. Digested proteome samples were resuspended in 80% acetonitrile and 1% TFA. Samples were then loaded onto equilibrated ZIC-HILIC columns. ZIC-HILIC columns were washed with 20 bed volumes of 80% acetonitrile and 1% TFA to remove nonglycosylated peptides and bound peptides eluted with 10 bed volumes of ultrapure water. Eluted peptides were dried by vacuum centrifugation and stored at −20°C.

TMT proteome analysis of swainsonine-treated murine T cells

TMT labeling was performed according to the manufacturer's instructions. Briefly, seven channels (130C, 131C, 129C, 128C, 127C, 126, and 127N) of a TMT10 kit (Thermo Fisher Scientific, Waltham, MA, USA) were used to label one biological sample. TMT reagents were allowed to warm to room temperature before being resuspended in 33 μl of acetonitrile and added to 100 μg of peptide aliquots resuspended in 100 mM of triethylammonium bicarbonate. Untreated biological replicates were labeled with the TMT labels 129C, 128C, 127C, and 126, while swainsonine-treated biological replicates were labeled with TMT labels 130C, 131C, and 127N. TMT labeling was allowed to proceed for 1 hour before being quenched with 5% hydroxylamine for 20 min. Samples were then combined, acidified with buffer A*, and desalted using 50-mg tC18 Sep-Pak columns, as described above. Eluted peptides were then dried down by vacuum centrifugation. The resulting TMT sample was then fractionated by Basic Reverse Phase C18 using the protocol of Batth and Olsen (*28*). The TMT sample was fractionated on a Gemini 5 μm NX-C18 250 mm by 4.6 mm column (Phenomenex, Torrance, CA, USA) using a 90-min gradient. The sample was resuspended in basic reverse-phase buffer A (5 mM ammonium hydroxide) and loaded onto the column at 1 ml/min for 3 min before introducing 2% basic reverse phase buffer B (5 mM ammonium hydroxide and 80% acetonitrile). The concentration of basic reverse phase buffer B was increased to 28% B over 45 min, then from 28% B to 40% B over 5 min, and then from 40% B to 100% B over 5 min. The composition was held at 100% B for 2 min and then dropped to 2% B over 10 min before being held at 2% B for another 10 min. One-minute

fractions were collected from 3 to 62 min and then dried down by vacuum centrifugation. The resulting 60 fractions were reduced to 12 fractions by combining every 12th fraction together. The combined fractions were dried by vacuum centrifugation then stored at −20°C until analysis by liquid chromatography–mass spectrometry (LC-MS).

L-Pha–based proximity labeling

For biotinylation of EL4 cells using L-Pha, 3×10^6 cells were incubated on ice for 1 hour with 66 μl of PBS containing biotinylated L-Pha (2.5 μg/ml) and 1% fetal calf serum (FCS). Nonbound L-Pha was washed away (×5, 1 ml) with 1% FCS in PBS; then, cells were resuspended in 66 μl of streptavidin-HRP at 5 μg/ml and 1% FCS in PBS and incubated on ice for 30 min and then at room temperature for 10 min. A total of 1.5 μl of tyramide-SS-biotin (Sigma-Aldrich) and 7 μl of 0.3% H_2O_2 solution were added and incubated at room temperature for 5 min. The reaction was terminated by washing the cells five times using ice-cold TBS and then snap-frozen in liquid nitrogen before enrichment of biotinylated proteins as described below.

Surface labeling of proteins with amino-oxy-biotin

Cells were washed (2×10^7) twice in ice-cold PBS followed by oxidized/biotinylated surface sialic acid residues with 1 mM sodium meta-periodate (Thermo Fisher Scientific, catalog no. 20504), 200 μM amino-oxy-biotin (Biotium, catalog no. 90113), and 10 mM aniline (Sigma-Aldrich, catalog no. 51788) in PBS for 1 hour in the dark at 4°C. The reaction was quenched by addition of glycerol to a final concentration of 1 mM. Cells were washed with cold PBS/5% FCS followed by another wash with cold PBS. Cells were then spun down (400*g*/4°C) and resuspended in lysis buffer [1% Triton X-100, 150 mM NaCl, 5 mM iodoacetamide, 10 mM tris-HCl (pH 7.6), and 1× protease inhibitor] for 30 min on ice. Nuclei were removed by centrifugation at 2800*g*, removal of the supernatant, and centrifugation again at 16000*g*. Cells were then snap-frozen before subsequent purification.

Label-free proteome analysis of biotinylated surface proteins

Enrichment of biotinylated proteins was undertaken as previously described (*29*). Briefly, snap-frozen cells were resuspended in 1 ml of ice-cold radioimmunoprecipitation assay (RIPA) lysis buffer [50 mM tris-HCl (pH 7.5), 150 mM NaCl, 1% NP-40, 1 mM EDTA, 0.1% SDS, and 0.5% sodium deoxycholate], supplemented with the cOmplete Protease Inhibitor Cocktail (Roche) and 250 U of Benzonase (Sigma-Aldrich), and then allowed to thaw on ice for 1 hour with agitation. Cells were then sonicated on ice to ensure complete solubilization, and then, samples were clarified with centrifugation at 10,000*g* and 4°C for 30 min. Protein concentrations were determined by a bicinchoninic acid assay (Thermo Fisher Scientific), and 50 μg was collected to be used as an input control. Clarified samples were then added to RIPA-washed Streptavidin Sepharose (GE Healthcare) and tumbled for 3 hours at 4°C. Streptavidin Sepharose was then washed with 1 ml of RIPA buffer, 4× 1 ml of PBS, and 4× 1 ml of 100 mM ABC. Streptavidin Sepharose beads were then resuspended in 200 μl of ABC containing 2 μg of trypsin (Promega) and incubated overnight at 37°C. Input controls were prepared according to the SP3 cleanup approach (*30*) and digested overnight with trypsin (1:50, w/w). The digested pulldowns and input controls were acidified with formic acid to 0.5% and then concentrated using homemade C18 stage tips (*27*). Eluted peptides were then dried by vacuum centrifugation and stored at −20°C until analysis by LC-MS.

Reverse-phase LC–tandem MS

Samples were resuspended in buffer A* and separated using a two-column chromatography setup composed of a PepMap 100 C18 20 mm by 75 μm trap and a PepMap C18 500 mm by 75 μm analytical column (Thermo Fisher Scientific). Samples were concentrated onto the trap column at 5 μl/min for 5 min with buffer A [0.1% formic acid and 2% dimethyl sulfoxide (DMSO)] and then infused into either an Orbitrap Q Exactive Plus, an Orbitrap Fusion Eclipse Tribrid, or an Orbitrap Elite Mass Spectrometer (Thermo Fisher Scientific) at 300 nl/min via the analytical column using a Dionex UltiMate 3000 UPLC (Thermo Fisher Scientific). Analytical runs (125 min) were undertaken by altering the buffer composition from 2% buffer B (0.1% formic acid, 77.9% acetonitrile, and 2% DMSO) to 28% B over 90 min, then from 28% B to 40% B over 10 min, and then from 40% B to 100% B over 2 min. The composition was held at 100% B for 3 min and then dropped to 2% B over 5 min before being held at 2% B for another 15 min. LFQ proteomics analysis of MGAT1 lines was undertaken using an Orbitrap Elite Mass Spectrometer operated in a data-dependent mode automatically switching between the acquisition of a single Orbitrap MS scan (maximum injection time of 50 ms, AGC (automatic gain control) 1×10^6, and 120,000 resolution) and up to 20 ion-trap collision-induced dissociation tandem MS (CID MS/MS) events [normalized collision energy (NCE) 30%; maximal injection time of 50 ms and AGC 2×10^4]. Glycopeptide analysis was undertaken using an Orbitrap Elite Mass Spectrometer operated in a data-dependent mode automatically switching between the acquisition of a single Orbitrap MS scan (maximum injection time of 50 ms, AGC 1×10^6, and 60,000 resolution) and the collection of up to five precursors with each precursors subjected to three fragmentation scans: an Orbitrap HCD (higher-energy collision dissociation) MS/MS scan (NCE, 42%; maximal injection time of 200 ms, with an AGC 2×10^5 and a resolution of 15,000), an ion-trap CID MS/MS scan (NCE 30%; maximal injection time of 50 ms and AGC 2×10^4), and an ion-trap ETD (electron-transfer dissociation) MS/MS scan (100-ms reaction time; maximal injection time of 50 ms and AGC 2×10^4). LFQ proteomics analysis of LFQ surface biotinylation experiments was analyzed on a Q Exactive mass spectrometer operated in a data-dependent mode automatically switching between the acquisition of a single Orbitrap MS scan (maximum injection time of 50 ms, AGC 3×10^6, and 70,000 resolution) and up to 15 MS/MS events (using stepped NCE 25, 30, and 35; maximal injection time of 110 ms and AGC 2×10^5 with a resolution of 15,000). TMT experiments were analyzed on an Eclipse mass spectrometer operated in a data-dependent mode automatically switching between the acquisitions of an Orbitrap MS scan (120,000 resolution) every 3 s and Orbitrap HCD MS/MS scans of precursors (NCE 38%; maximal injection time of 80 ms, with an AGC of 300% and a resolution of 60,000). Glycan fragment ion (204.087, 366.1396, and 138.0545 mass/charge ratio) product-dependent MS/MS analysis (31) was used to trigger two additional scans of potential glycopeptides: an ion trap CID scan (NCE 35%; maximal injection time of 40 ms with an AGC of 200%) and a stepped collision energy HCD scan (using NCE 30, 40, and 48% with a maximal injection time of 200 ms with an AGC of 500% and a resolution of 60,000).

Proteomic data analysis—Quantitative proteomic analysis

LFQ- and TMT-based experiments were analyzed using MaxQuant (v1.6.3.4 or 1.5.3.30) (32). Searches were performed against the *Mus musculus* databases (UniProt accession: UP000000589) with carbamidomethylation of cysteine set as a fixed modification and the variable modifications, acetylation of protein N termini and oxidation of methionine (M). For TMT experiments, the TMT10 channels 130C, 131C, 129C, 128C, 127C, 126, and 127N were allowed and the modification of the N termini and lysine with TMT10 were included as fixed modifications. Searches were performed with trypsin cleavage specificity, allowing two missed cleavage events. The precursor mass tolerance was set to 20 parts per million (ppm) for the first search and 10 ppm for the main search, with a maximum false discovery rate of 1.0% set for protein and peptide identifications. For LFQ experiments, the "match between run" option (33) was enabled to improve the detection of peptides between samples. The resulting protein group outputs were processed within the Perseus (v1.4.0.6) (34) analysis environment to remove reverse matches and common protein contaminates before quantitative analysis. For LFQ experiments, missing values were imputed on the basis of the observed total peptide intensities with a range of 0.3σ and a downshift of 2.0σ. Samples were grouped, and Student's *t* test was used to assign *P* values; multiple hypothesis correction was undertaken using a Benjamini-Hochberg correction.

Proteomic data analysis—Glycopeptide identification

Raw data files were batch processed using Byonic v3.5.3 or 2.13.2 [Protein Metrics Inc. (35)]. Data were searched on a desktop with two 3.00-GHz Intel Xeon Gold 6148 processors, a 2-TB solid-state drive, and 128 GB of random-access memory using a maximum of 16 cores for a given search. For all searches, trypsin specificity was set and a maximum of two missed cleavage events was allowed. Carbamidomethyl was set as a fixed modification of cystine, while oxidation of methionine was included as a variable modification. The default Byonic N-linked glycan database, which is composed of 309 mammalian N-glycans, was used. A maximum mass precursor tolerance of 5 ppm was allowed, while a mass tolerance of up to 10 ppm was set for HCD fragments and 20 ppm for EThcD (electron-transfer/higher-energy collision dissociation) fragments. To ensure high data quality, technical replicates were combined using R (www.r-project.org/) and only glycopeptides with a Byonic score >150 were used for further analysis. This score cutoff is in line with previous reports highlighting that score thresholds greater than at least 150 are required for robust glycopeptide assignments with Byonic (36).

Proteomic data availability

All proteomics data have been deposited to the ProteomeXchange Consortium via the proteomics identifications database (PRIDE) (37) partner repository with the dataset identifier PXD020685. Data can be accessed using the username: reviewer47711@ebi.ac.uk and password: nIxeYuQz.

REFERENCES AND NOTES

1. M. Stoeckius, C. Hafemeister, W. Stephenson, B. Houck-Loomis, P. K. Chattopadhyay, H. Swerdlow, R. Satija, P. Smibert, Simultaneous epitope and transcriptome measurement in single cells. *Nat. Methods* **14**, 865–868 (2017).

2. V. M. Peterson, K. X. Zhang, N. Kumar, J. Wong, L. Li, D. C. Wilson, R. Moore, T. K. McClanahan, S. Sadekova, J. A. Klappenbach, Multiplexed quantification of proteins and transcripts in single cells. *Nat. Biotechnol.* **35**, 936–939 (2017).

3. D. Lahnemann, J. Köster, E. Szczurek, D. J. McCarthy, S. C. Hicks, M. D. Robinson, C. A. Vallejos, K. R. Campbell, N. Beerenwinkel, A. Mahfouz, L. Pinello, P. Skums, A. Stamatakis, C. S. Attolini, S. Aparicio, J. Baaijens, M. Balvert, B. Barbanson, A. Cappuccio, G. Corleone, B. E. Dutilh, M. Florescu, V. Guryev, R. Holmer, K. Jahn, T. J. Lobo, E. M. Keizer, I. Khatri, S. M. Kielbasa, J. O. Korbel, A. M. Kozlov, T. H. Kuo, B. P. F. Lelieveldt, I. I. Mandoiu, J. C. Marioni, T. Marschall, F. Molder, A. Niknejad, L. Raczkowski, M. Reinders, J. Ridder, A. E. Saliba, S. Somarakis, O. Stegle, F. J. Theis, H. Yang, A. Zelikovsky, A. C. McHardy, B. J. Raphael, S. P. Shah, A. Schönhuth, Eleven grand challenges in single-cell data science. *Genome Biol.* **21**, 31 (2020).

4. E. Z. Macosko, A. Basu, R. Satija, J. Nemesh, K. Shekhar, M. Goldman, I. Tirosh, A. R. Bialas, N. Kamitaki, E. M. Martersteck, J. J. Trombetta, D. A. Weitz, J. R. Sanes, A. K. Shalek, A. Regev, S. A. McCarroll, Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. *Cell* **161**, 1202–1214 (2015).

5. C. Buccitelli, M. Selbach, mRNAs, proteins and the emerging principles of gene expression control. *Nat. Rev. Genet.* **21**, 630–644 (2020).

6. K. W. Moremen, M. Tiemeyer, A. V. Nairn, Vertebrate protein glycosylation: Diversity, synthesis and function. *Nat. Rev. Mol. Cell Biol.* **13**, 448–462 (2012).

7. R. Kornfeld, S. Kornfeld, Assembly of asparagine-linked oligosaccharides. *Annu. Rev. Biochem.* **54**, 631–664 (1985).

8. M. S. Pereira, I. Alves, M. Vicente, A. Campar, M. C. Silva, N. A. Padrao, V. Pinto, A. Fernandes, A. M. Dias, S. S. Pinho, Glycans as key checkpoints of T cell activity and function. *Front. Immunol.* **9**, 2754 (2018).

9. J. D. Marth, P. K. Grewal, Mammalian glycosylation in immunity. *Nat. Rev. Immunol.* **8**, 874–887 (2008).

10. L. K. Smith, G. M. Boukhaled, S. A. Condotta, S. Mazouz, J. J. Guthmiller, R. Vijay, N. S. Butler, J. Bruneau, N. H. Shoukry, C. M. Krawczyk, M. J. Richer, Interleukin-10 directly inhibits CD8+ T cell function by enhancing N-glycan branching to decrease antigen sensitivity. *Immunity* **48**, 299–312.e5 (2018).

11. M. Demetriou, M. Granovsky, S. Quaggin, J. W. Dennis, Negative regulation of T-cell activation and autoimmunity by Mgat5 N-glycosylation. *Nature* **409**, 733–739 (2001).

12. A. M. Dias, A. Correia, M. S. Pereira, C. R. Almeida, I. Alves, V. Pinto, T. A. Catarino, N. Mendes, M. Leander, M. T. Oliva-Teles, L. Maia, C. Delerue-Matos, N. Taniguchi, M. Lima, I. Pedroto, R. Marcos-Pinto, P. Lago, C. A. Reis, M. Vilanova, S. S. Pinho, Metabolic control of T cell immune response through glycans in inflammatory bowel disease. *Proc. Natl. Acad. Sci. U.S.A.* **115**, E4651–E4660 (2018).

13. E. De Bousser, L. Meuris, N. Callewaert, N. Festjens, Human T cell glycosylation and implications on immune therapy for cancer. *Hum. Vaccin. Immunother.*, 2374–2388 (2020).

14. L. Araujo, P. Khim, H. Mkhikian, C. L. Mortales, M. Demetriou, Glycolysis and glutaminolysis cooperatively control T cell function by limiting metabolite supply to N-glycosylation. *eLife* **6**, e21330 (2017).

15. J. P. Ribeiro, L. K. Mahal, Dot by dot: Analyzing the glycome using lectin microarrays. *Curr. Opin. Chem. Biol.* **17**, 827–831 (2013).

16. H. Ghazarian, B. Idoni, S. B. Oppenheimer, A glycobiology review: Carbohydrates, lectins and implications in cancer therapeutics. *Acta Histochem.* **113**, 236–247 (2011).

17. N. Sharon, H. Lis, History of lectins: From hemagglutinins to biological recognition molecules. *Glycobiology* **14**, 53R–62R (2004).

18. M. Aebi, N-linked protein glycosylation in the ER. *Biochim. Biophys. Acta* **1833**, 2430–2437 (2013).

19. J. J. Havel, D. Chowell, T. A. Chan, The evolving landscape of biomarkers for checkpoint inhibitor immunotherapy. *Nat. Rev. Cancer* **19**, 133–150 (2019).

20. G. Wu, M. Nagala, P. R. Crocker, Identification of lectin counter-receptors on cell membranes by proximity labeling. *Glycobiology* **27**, 800–805 (2017).

21. S. L. Wolock, R. Lopez, A. M. Klein, Scrublet: Computational identification of cell doublets in single-cell transcriptomic data. *Cell Syst* **8**, 281–291.e9 (2019).

22. S. Aibar, C. B. Gonzalez-Blas, T. Moerman, V. A. Huynh-Thu, H. Imrichova, G. Hulselmans, F. Rambow, J. C. Marine, P. Geurts, J. Aerts, J. van den Oord, Z. K. Atak, J. Wouters, S. Aerts, SCENIC: Single-cell regulatory network inference and clustering. *Nat. Methods* **14**, 1083–1086 (2017).

23. T. Stuart, A. Butler, P. Hoffman, C. Hafemeister, E. Papalexi, W. M. Mauck III, Y. Hao, M. Stoeckius, P. Smibert, R. Satija, Comprehensive integration of single-cell data. *Cell* **177**, 1888–1902.e21 (2019).

24. A. J. Freeman, S. J. Vervoort, K. M. Ramsbottom, M. J. Kelly, J. Michie, L. Pijpers, R. W. Johnstone, C. J. Kearney, J. Oliaro, Natural killer cells suppress T cell-associated tumor immune evasion. *Cell Rep.* **28**, 2784–2794.e5 (2019).

25. S. J. Humphrey, S. B. Azimifar, M. Mann, High-throughput phosphoproteomics reveals in vivo insulin signaling dynamics. *Nat. Biotechnol.* **33**, 990–995 (2015).

26. S. Mysling, G. Palmisano, P. Hojrup, M. Thaysen-Andersen, Utilizing ion-pairing hydrophilic interaction chromatography solid phase extraction for efficient glycopeptide enrichment in glycoproteomics. *Anal. Chem.* **82**, 5598–5609 (2010).

27. J. Rappsilber, M. Mann, Y. Ishihama, Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips. *Nat. Protoc.* **2**, 1896–1906 (2007).

28. T. S. Batth, J. V. Olsen, Offline high pH reversed-phase peptide fractionation for deep phosphoproteome coverage. *Methods Mol. Biol.* **1355**, 179–192 (2016).

29. E. Coyaud, M. Mis, E. M. Laurent, W. H. Dunham, A. L. Couzens, M. Robitaille, A. C. Gingras, S. Angers, B. Raught, BioID-based identification of Skp Cullin F-box (SCF)β-TrCP1/2E3 ligase substrates. *Mol. Cell. Proteomics* **14**, 1781–1795 (2015).

30. C. S. Hughes, S. Moggridge, T. Muller, P. H. Sorensen, G. B. Morin, J. Krijgsveld, Single-pot, solid-phase-enhanced sample preparation for proteomics experiments. *Nat. Protoc.* **14**, 68–85 (2019).

31. J. Saba, S. Dutta, E. Hemenway, R. Viner, Increasing the productivity of glycopeptides analysis by using higher-energy collision dissociation-accurate mass-product-dependent electron transfer dissociation. *Int. J. Proteomics* **2012**, 560391 (2012).

32. J. Cox, M. Mann, MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat. Biotechnol.* **26**, 1367–1372 (2008).

33. J. Cox, M. Y. Hein, C. A. Luber, I. Paron, N. Nagaraj, M. Mann, Accurate proteome-wide label-free quantification by delayed normalization and maximal peptide ratio extraction, termed MaxLFQ. *Mol. Cell. Proteomics* **13**, 2513–2526 (2014).

34. S. Tyanova, T. Temu, P. Sinitcyn, A. Carlson, M. Y. Hein, T. Geiger, M. Mann, J. Cox, The Perseus computational platform for comprehensive analysis of (prote)omics data. *Nat. Methods* **13**, 731–740 (2016).

35. M. Bern, Y. J. Kil, C. Becker, Byonic: Advanced peptide and protein identification software. *Curr. Protoc. Bioinformatics* **Chapter 13**, Unit13.20 (2012).

36. L. Y. Lee, E. S. Moh, B. L. Parker, M. Bern, N. H. Packer, M. Thaysen-Andersen, Toward automated N-glycopeptide identification in glycoproteomics. *J. Proteome Res.* **15**, 3904–3915 (2016).

37. Y. Perez-Riverol, A. Csordas, J. Bai, M. Bernal-Llinares, S. Hewapathirana, D. J. Kundu, A. Inuganti, J. Griss, G. Mayer, M. Eisenacher, E. Perez, J. Uszkoreit, J. Pfeuffer, T. Sachsenberg, S. Yilmaz, S. Tiwary, J. Cox, E. Audain, M. Walzer, A. F. Jarnuczak, T. Ternent, A. Brazma, J. A. Vizcaino, The PRIDE database and related tools and resources in 2019: Improving support for quantification data. *Nucleic Acids Res.* **47**, D442–D450 (2019).

Citation: C. J. Kearney, S. J. Vervoort, K. M. Ramsbottom, I. Todorovski, E. J. Lelliott, M. Zethoven, L. Pijpers, B. P. Martin, T. Semple, L. Martelotto, J. A. Trapani, I. A. Parish, N. E. Scott, J. Oliaro, R. W. Johnstone, SUGAR-seq enables simultaneous detection of glycans, epitopes, and the transcriptome in single cells. *Sci. Adv.* **7**, eabe3610 (2021).