Functional proteogenomics reveals biomarkers and therapeutic targets in lymphomas

Delphine C. M. Rolland[a,1], Venkatesha Basrur[b,1], Yoon-Kyung Jeon[c,1], Carla McNeil-Schwalm[d], Damian Fermin[b], Kevin P. Conlon[b], Yeqiao Zhou[a], Samuel Y. Ng[e], Chih-Chiang Tsou[b], Noah A. Brown[b], Dafydd G. Thomas[b], Nathanael G. Bailey[b], Gilbert S. Omenn[f,g,h], Alexey I. Nesvizhskii[b,h], David E. Root[i], David M. Weinstock[e,i], Robert B. Faryabi[a,j], Megan S. Lim[a,2], and Kojo S. J. Elenitoba-Johnson[a,j,2]

[a]Department of Pathology and Laboratory Medicine, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA 19104; [b]Department of Pathology, University of Michigan Medical School, Ann Arbor, MI 48109; [c]Department of Pathology, Seoul National University Hospital, Seoul National University College of Medicine, Seoul, South Korea; [d]Department of Pediatrics, University of Michigan Medical School, Ann Arbor, MI 48109; [e]Department of Medical Oncology, Dana–Farber Cancer Institute, Harvard Medical School, Boston, MA 02215; [f]Department of Internal Medicine, University of Michigan Medical School, Ann Arbor, MI 48109; [g]Department of Human Genetics, University of Michigan Medical School, Ann Arbor, MI 48109; [h]Department of Computational Medicine and Bioinformatics, University of Michigan Medical School, Ann Arbor, MI 48109; [i]The Broad Institute of Massachusetts Institute of Technology and Harvard Medical School, Boston, MA 02142; and [j]Center for Personalized Diagnostics, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA 19104

Identification of biomarkers and therapeutic targets is a critical goal of precision medicine. N-glycoproteins are a particularly attractive class of proteins that constitute potential cancer biomarkers and therapeutic targets for small molecules, antibodies, and cellular therapies. Using mass spectrometry (MS), we generated a compendium of 1,091 N-glycoproteins (from 40 human primary lymphomas and cell lines). Hierarchical clustering revealed distinct subtype signatures that included several subtype-specific biomarkers. Orthogonal immunological studies in 671 primary lymphoma tissue biopsies and 32 lymphoma-derived cell lines corroborated MS data. In anaplastic lymphoma kinase-positive (ALK[+]) anaplastic large cell lymphoma (ALCL), integration of N-glycoproteomics and transcriptome sequencing revealed an ALK-regulated cytokine/receptor signaling network, including vulnerabilities corroborated by a genome-wide clustered regularly interspaced short palindromic screen. Functional targeting of IL-31 receptor β, an ALCL-enriched and ALK-regulated N-glycoprotein in this network, abrogated ALK[+]ALCL growth in vitro and in vivo. Our results highlight the utility of functional proteogenomic approaches for discovery of cancer biomarkers and therapeutic targets.

proteomics | RNA-seq | CRISPR screen | biomarkers | lymphoma

The discovery of biologically relevant biomarkers and therapeutic targets is a critical goal of precision medicine. Integrative strategies combining multiple large-scale analyses, such as genomics, transcriptomics, and proteomics, offer complementary opportunities for the elucidation of novel pathogenic insights and discovery of functional modules and networks that are dysregulated in disease. Although the genomic aberrations associated with major types of primary cancer are being extensively characterized, the proteomic signatures of many cancers, including lymphomas, are unknown.

Protein glycosylation is one of the most common posttranslational modifications (PTMs) and plays important roles in many biological processes. N-glycosylation occurs on the amine group of asparagine residues (1, 2) and facilitates protein trafficking to membranes and secretion into the extracellular environment (3). N-glycoproteins are already used as diagnostic/prognostic biomarkers in clinical practice. The N-glycoproteins include many cluster of differentiation (CD) proteins that define hematopoietic subpopulations, as well as hematopoietic neoplasms (4). Many of these proteins have been found to represent useful targets for therapeutic antibodies, immunotoxins, and chimeric antigen receptor T cells (5–8). Thus, large-scale characterization of N-glycoproteins is appealing for discovering biomarkers and therapeutic targets.

To identify novel biomarkers and therapeutic targets in lymphomas, we performed agnostic mass spectrometry (MS)-based profiling of N-glycoproteins of 13 distinct subtypes of lymphomas. Orthogonal immunophenotypic studies in primary lymphoma samples corroborated MS-based results and revealed several

biomarker candidates. We focused on anaplastic lymphoma kinase-positive (ALK[+]) anaplastic large cell lymphoma (ALCL) to obtain a more complete view of the "regulome" controlled by oncogenically activated ALK. To this end, we integrated RNA sequencing (RNA-seq) with N-glycoproteomic data to reveal ALK-regulated functional signaling networks. We functionally validated the expression and oncogenic role of selected components as being biologically relevant therapeutic targets in ALK[+]ALCL. Recently, large-scale analysis of cancer susceptibility genes has been greatly facilitated by using the bacterial clustered regularly interspaced short palindromic (CRISPR) system in combination with genome-wide, single-guide RNA (sgRNA) libraries to identify genes essential for cancer cell survival (9). Accordingly, we integrated the results of a genome-wide, CRISPR-Cas9–mediated vulnerability screen, which confirmed several susceptible therapeutic targets, including proteins in the signaling axes identified by proteogenomic integration. Our findings demonstrate the advantages of integration of multiomics data for the discovery of novel cancer biomarkers.

Significance

An important goal in precision oncology is the identification of biomarkers and therapeutic targets. We identified and annotated a compendium of N-glycoproteins from diverse human lymphoid neoplasia, an attractive class of proteins with potential to serve as cancer biomarkers and therapeutic targets. In anaplastic lymphoma kinase-positive (ALK[+]) anaplastic large cell lymphoma (ALCL), integration of N-glycoproteomics and transcriptome sequencing revealed an underappreciated and targetable ALK-regulated cytokine/receptor signaling network highlighting the utility of functional proteogenomics for discovery of cancer biomarkers and therapeutic targets.

MEDICAL SCIENCES

Results

Unbiased Analysis of N-Glycoproteomes of Human Lymphomas. We performed solid-phase extraction of glycoproteins (SPEG) (10), followed by identification of PNGase-F deglycosylated peptides using liquid chromatography (LC)-tandem MS (MS/MS) (Fig. S1*A*). Label-free spectral counting (11) was used to quantify the relative abundance of N-glycoproteins in 32 human cell lines representing 13 subtypes of lymphoma (Figs. S1*B* and S2*A*). The data have been deposited to the ProteomeXchange Consortium (proteomecentral.proteomexchange.org/cgi/GetDataset) via the PRIDE (PRoteomics IDEntifications database) partner repository (12) with the dataset identifier PXD003469. In total, 1,091 unique N-glycoproteins were identified. The list of N-glycoproteins and their observed spectral counts is provided in Dataset S1. Overall, 69.5% of the identified N-glycoproteins were assigned to membrane compartments, of which 52.3% could be localized to the plasma/integral to membrane fraction. The endoplasmic reticulum, Golgi, and lysosome membranes were also highly represented (20.3%), whereas 10.1% were annotated as secreted proteins (Fig. S2*B*). Using the SOSUI transmembrane domain (TMD) prediction algorithm (13), we noted that 64.6% of the membrane proteins contained one to two TMDs. Using the MotifX algorithm (14), we confirmed that the most significantly enriched motif in all N-glycoproteins was the classical N-glycosylation sequence: N-!P-[S/T]. Another significantly enriched motif was a subset of the N-!P-S motif, where Y at the −1 position was overrepresented. Finally, gene set enrichment analysis identified proteins involved in "cell adhesion," "signal transduction," "immune response," and "receptor activity" as significantly represented in our dataset (Fig. S2 *C* and *D*).

N-Glycoproteomic Profiles Classify Lymphomas According to Lineage, Cell of Origin, and Subtype. Unsupervised clustering using 751 unique N-glycoproteins that met stipulated criteria (*Methods*) discriminated T/NK (natural killer)-cell neoplasia and B-cell lymphomas (Fig. 1*A*). Within the T/NK-cell group, all seven ALCL cell lines, all three NK-cell lymphoma cell lines and the two cutaneous T-cell lymphoma cells lines formed distinct clades. Similarly, in the B-cell lymphoma group, classical Hodgkin lymphoma (cHL) and non-Hodgkin lymphoma (NHL) cell lines formed two distinct clades. Among the B-cell NHLs, cell lines representative of pre-germinal (pre-GC) center lymphoid proliferation [mantle-cell lymphoma (MCL)] grouped together, whereas GC-derived cell lines [Burkitt lymphoma (BL), transformed follicular lymphoma (t-FL), and diffuse large B-cell lymphoma (DLBCL)] formed a distinctive subgroup. The data show that pre-GC and GC-derived B-lymphomas have distinct N-glycoproteomic profiles.

To assess the diagnostic validity of this approach, we performed N-glycoproteomic profiling on eight primary lymphoma samples in a blinded fashion and evaluated the correlation of their N-glycoprotein profile with the cell lines. The glycoproteomic profiles of clinical samples of patients (P)1–P4 highly correlated with MCL cell lines, whereas the profiles of clinical samples P5–P8 matched with GC-derived NHL cell lines (Fig. 1*B*). These observations were in complete agreement with the clinical diagnosis rendered for these samples; clinical samples P1–P4 represented well-characterized cases of MCL, and clinical samples P5–P8 represented well-characterized cases of FL (Table S1). These data indicate that N-glycoproteomic profiles can accurately cluster primary lymphoma specimens into appropriate diagnostic categories.

N-Glycoproteomic Analysis Accurately Identifies All Clinically Relevant CD Proteins. Of the 417 CD proteins listed by the Human Cell Differentiation Molecules consortium (www.hcdm.org/), we identified 194 (46.5%) CD proteins in our N-glycoproteomic dataset (Dataset S1). We identified between 52 and 102 CD proteins per WHO subtype (average of 79) (Fig. S3*A*). Importantly, we detected virtually all CD proteins currently used for diagnostic evaluation of lymphomas. Pan–B-cell markers (CD19, CD22, CD79a, and CD79b) were appropriately identified in cell lines representative of MCL, t-FL, DLBCL, primary mediastinal
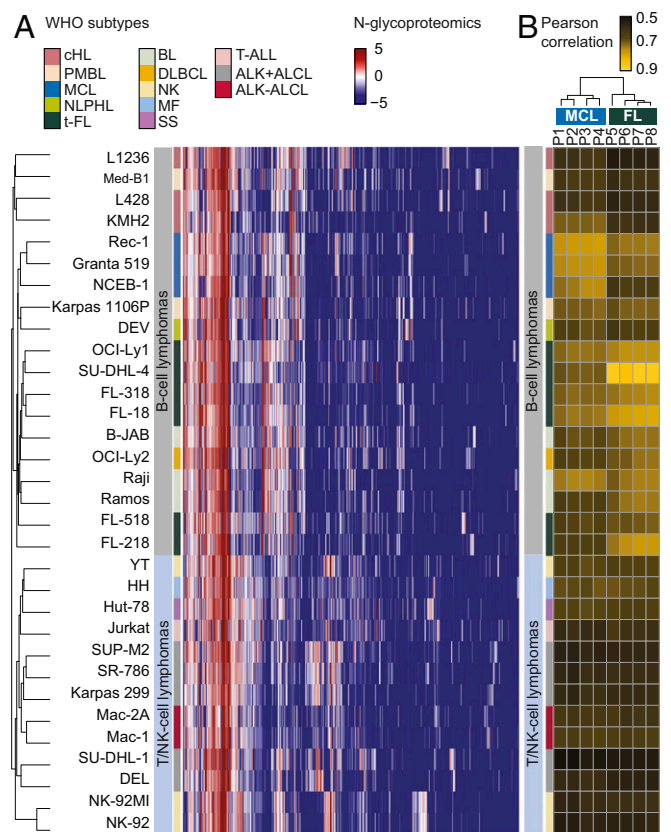

Fig. 1. N-glycoproteomic profiles classify lymphomas according to lineage, cell of origin, and WHO subtype. (*A*) Thirty-two cell lines included in this study are accurately classified according to their lineage (B or T/NK cells), cell of origin (pre-GC– and GC-derived B cells), and WHO subtypes. (*B*) Clustering of eight blinded primary B-cell lymphoma samples based on their Pearson's correlations with the 32 cell lines accurately designated the profiles of clinical samples P1 to P4, which are highly correlated to MCL cell lines, whereas clinical samples P5 to P8 are highly correlated to t-FL cell lines. NLPHL, nodular lymphocyte predominant Hodgkin lymphoma; MF, mycosis fungoides; SS, Sézary syndrome; T-ALL, T-lymphoblastic leukemia/lymphoma.

B-cell lymphoma (PMBL) and nodular lymphocyte predominant Hodgkin lymphoma (NLPHL), but not in cell lines with T/NK-cell origin or cell lines representative of cHL, which typically lack expression of these markers. In addition, CD10 was identified in BL-, DLBCL-, PMBL-, and t-FL–derived cell lines but not in any other B-cell neoplasia or in any cell lines from T/NK-cell origin (Dataset S1). Furthermore, CD30 was also detected as expected in cell lines representative of cHL, PMBL, ALCL (ALK$^+$ and ALK$^-$), Sézary syndrome, mycosis fungoides, and aggressive NK-cell leukemia (15) (Fig. S3*B*). Moreover, the spectral counts for CD30 were consistent with the relative intensities observed by Western blot analysis of cHL, PMBL, and ALK$^+$ALCL cell lines (15). As frequently observed in T/NK-cell neoplasia, the pan–T-cell markers (CD2, CD3, CD4, CD5, and CD7) were frequently singly or combinatorially undetected in T/NK-cell–derived lymphoma cell lines (16). CD56 was detected in two of the three aggressive NK-cell lymphoma cell lines (Dataset S1). Correspondingly, the primary lymphoma samples uniformly showed expression of the pan–B-cell markers CD79a and CD79b. Notably, CD10 was detected exclusively in the FL samples and not in the MCL samples, in keeping with the expected immunophenotypic profiles of the two B-cell lymphoma subcategories. Taken together, our analysis of N-glycoproteins represents a compendium of CD proteins expressed in lymphomas. The expression patterns of the vast majority of these proteins have not previously been described in the diverse lymphoma subtypes. Their

identification in this study offers opportunities for their exploration as diagnostic biomarkers and therapeutic targets.

Immunophenotypic Validation of Selected N-Glycoproteins in Primary Clinical Samples.

For further orthogonal validation, we selected three proteins whose expression is underexplored in mature B-cell lymphomas and were observed by MS to be differentially expressed between t-FL–, BL-, and MCL-derived cell lines: CD44, Paraoxonase 2 (PON2), and CD276 (Fig. S4A). CD44 was identified in all MCL cell lines but absent or largely negative in BL and t-FL cell lines. These observations were confirmed by Western blotting and flow cytometry analyses (combined Fisher's P value < 1e-10; Fig. S4 A and B). PON2 was also identified consistently in all MCL cell lines, although its expression was more limited in other lymphomas. Western blot analysis and spectral counts of PON2 demonstrated expression in MCL among the B-cell lymphomas and in ALCL cell lines (Fig. S4A). PON2 protein expression in lymphomas has not received notable attention in the literature, although differential expression of PON2 mRNA has been suggested in a single report (17). We further analyzed the expression of CD44 and PON2 by immunohistochemistry on lymphoma tissue microarrays containing 34 clinical biopsy specimens and revealed that 91.2% of MCL expressed CD44 and 58.8% of MCL expressed PON2 (Fig. S4C). As shown in Fig. S4A, CD276 was expressed in a subset of MCL and ALK⁺ALCL cell lines but not in GC-derived lymphomas, such as t-FL and BL. We observed significant concordance (combined Fisher's P value < 1e-9) between spectral counts, Western blot analysis, and flow cytometric analysis for CD276 (Fig. S4D). This observation was corroborated by the immunohistochemistry of primary lymphoma tissue biopsies ($n = 644$), which showed that 35.3% of MCL and 18.8% of ALK⁺ALCL expressed CD276, whereas only rare cases of other lymphomas demonstrated CD276 expression (P < 1e-4; Fig. S4E).

Taken together, our MS-based profiling of N-glycoproteins identified several potential biomarkers of which a subset was further orthogonally validated in a large cohort of primary lymphoma tissue biopsies. Some of these potential biomarkers demonstrated highly selective expression in some subtypes of lymphoma. Using ALK⁺ALCL as a model, we investigated whether restricted expression may be linked to the pathogenesis of these subtypes, and performed functional studies to assess the candidacy of a selected protein as a vulnerability target.

Integrative Transcriptomics and Proteomics Reveal ALK-Activated Cytokine Receptor Network in ALCL.

To identify novel biologically relevant N-glycoproteins that play critical roles in the pathogenesis of ALK⁺ALCL, we used an integrative N-glycoproteomics and transcriptomics RNA-seq strategy. N-glycoproteomic analysis of ALK⁺ALCL identified several proteins with "cytokine receptor" activity, including inflammatory cytokine receptors [IL-1 receptor 1 (IL-1R1), IL-1R2, and IL-1RAP), Th1 cytokine receptors (IL-2Rα and IL-18R), Th2 cytokine receptor (IL-4Rα), and Th17 cytokine/cytokine receptors (IL-17, IL-17Rα, IL-22, IL-6Rα, IL-6Rβ, IL-31Rα, and IL-31Rβ) (Dataset S1)].

To determine whether the expression of the cytokine receptor signaling network is regulated by ALK activity, we performed N-glycoproteomic profiling of two ALK⁺ALCL cell lines (Karpas 299 and SU-DHL-1) with or without treatment for 9 h with CEP-26939 (300 nM), an ALK inhibitor (ALKi). We observed that 65 and 90 N-glycoproteins exhibited decreased expression in Karpas 299 and SU-DHL-1, respectively, upon treatment, whereas 11 N-glycoproteins decreased in both cell lines, including interleukin receptors, such as IL-31Rβ (Fig. S5). These results indicate that expression of these glycoproteins is regulated by ALK activity.

We also performed RNA-seq analysis of these two ALK⁺ALCL cell lines treated with crizotinib (100 nM), a US Food and Drug Administration-approved ALKi (18). The RNA-seq data were deposited in the Gene Expression Omnibus (19) with the accession number GSE81301. We observed that 1,244 transcripts were altered by ALK inhibition in both cell lines, with 400 transcripts

being induced and 844 being repressed (Fig. S6A). In either cell line, the ALK-induced transcripts were significantly enriched for members of cytokine/receptor-STAT signaling pathways (Fig. S6 B and C). Furthermore, gene set enrichment analysis (GSEA) (20) revealed that transcripts involved in IL-2–STAT5, IL-6–STAT3 and TNF-α signaling pathways were significantly down-regulated after ALK inhibition, suggesting up-regulation of these pathways in ALK⁺ALCL (Fig. S6 D and E).

Integration of N-glycoproteomic and transcriptomic datasets revealed both corroborative and complementary information about the pathobiology of ALK⁺ALCL. The Spearman's correlations between the steady-state mRNA and N-glycoprotein abundance were 0.50 and 0.49 in Karpas 299 and SU-DHL-1, respectively, which is comparable to previous reports (21) and suggests that RNA-seq and N-glycoproteomic dataset analyses may yield complementary results. Approximately 13% of the gene products were only measured at the N-glycoprotein level (Dataset S2), whereas more than 34% of the CD genes measured by either method were not detected by N-glycoproteomics, suggesting the importance of multiomic approaches for comprehensive analyses of cellular processes (Dataset S2). Despite these differences, in both Karpas 299 and SU-DHL-1, the N-glycoproteins induced by ALK were enriched in the ALK-regulated transcripts, indicating concordance between transcriptional and N-glycoproteomic levels [false discovery rate (FDR) < 1e-2; Fig. 2 A and B].

Integrative unsupervised clustering of N-glycoproteomic and transcriptomic datasets of both cell lines revealed 36 gene products that were concordantly decreased as a result of ALK inhibition in transcriptomic and glycoproteomic data of both cell lines (Fig. 2C). Several of these gene products are members of cytokine/receptor families, such as IL-1R1, IL-1R2, IL-2Rα, IL-4R, IL-18R1, and IL-31Rβ and STAT signaling networks (IL-6–STAT3 and IL-2–STAT5) (Fig. 2D). Reconstruction of the protein interaction network of these cytokine receptors revealed that several key effectors, such as STAT3, STAT1, JAK1/2, and STAT5B, are also induced by ALK in ALCL (Fig. 2E).

Identification of Selective Expression of IL-31Rβ in ALK⁺ALCL, Regulation by ALK Activity, and Contribution to ALK–Mediated Oncogenesis.

To investigate the validity of novel biomarkers identified by N-glycoproteomic and transcriptomic studies further, we investigated IL-31Rβ, a cytokine receptor we noted to be exclusively expressed in ALK⁺ALCL cell lines by N-glycoproteomics (Dataset S1) and regulated by ALK through transcription. Western blotting and flow cytometry results confirmed the selective expression of IL-31Rβ in all ALK⁺ALCL cell lines (Fig. 3A and Fig. S7A). Immunohistochemistry on 56 ALCL biopsies (ALK⁺ and ALK⁻) revealed that 100% of ALK⁺ALCL expressed IL-31Rβ, whereas only 40% of ALK⁻ALCL was positive (χ^2 = 20.6, P < 1e-3; Fig. 3B). Given the selectivity of its expression in ALK⁺ALCL, we sought to assess its regulation by ALK. Pharmacological inhibition of ALK by two inhibitors (crizotinib and CEP-26939) decreased IL-31Rβ protein expression in a dose- and time-dependent manner (Fig. 3C and Fig. S7 B and C). To substantiate the effect of ALK activity on IL-31Rβ expression further, we ectopically expressed nucleophosmin–anaplastic lymphoma kinase (NPM-ALK) and its kinase-dead mutant NPM-ALK K210R into HeLa cells. As shown in Fig. S7D, NPM-ALK, but not the kinase-dead mutant K210R, induced IL-31Rβ expression, confirming the regulation of IL-31Rβ expression by ALK. Using quantitative real-time PCR, we confirmed that ALK regulates IL-31Rβ expression at the transcriptional level (Fig. S7 E and F). We also demonstrated that the transcriptional expression of oncostatin M (OSM), the ligand of IL-31Rβ, is dependent on ALK activity (Fig. 2E and Fig. S6A) and corroborated at the protein level by Western blotting (Fig. 3C). Interestingly, OSM partially rescued the phosphorylation of STAT3, indicative of its activation when ALK was inhibited (Fig. S7G). This observation suggests that the OSM/IL-31Rβ/STAT3 signaling axis (22) might function as a positive regulatory loop to maintain consistent STAT3 activation in ALK⁺ALCL (23) (Fig. S7H).
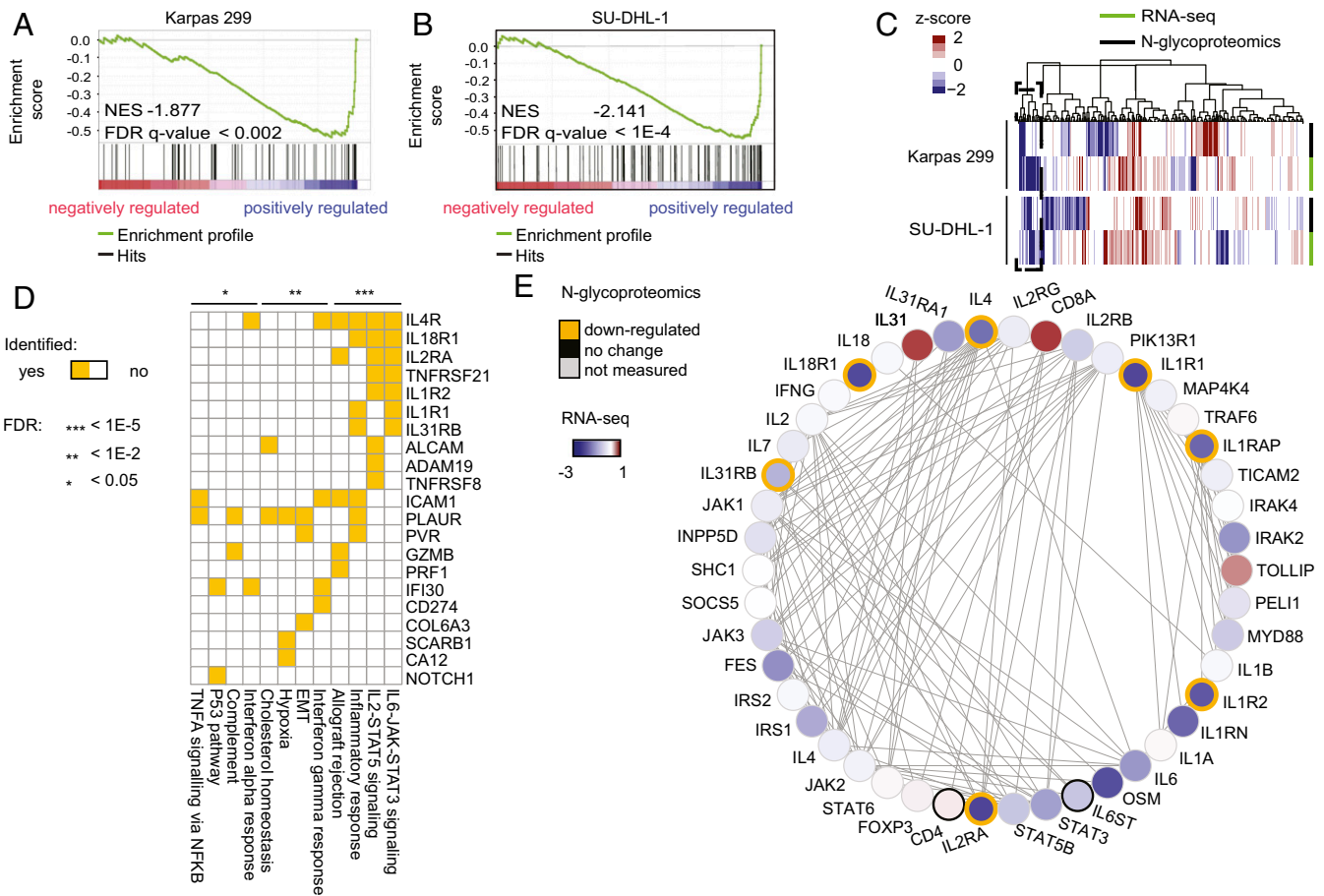
Rolland et al.

MEDICAL SCIENCES

Fig. 2. Integration of N-glycoproteome and transcriptome. (*A* and *B*) GSEA shows enrichment of the ALK$^+$ regulated N-glycoproteins in genes whose transcription decreases after ALK inhibition. NES, normalized enrichment score. (*C*) Integrative unsupervised clustering identifies 36 genes (dashed-line box) with a concomitant decrease in transcriptional and protein expression levels after ALK inhibition in both cell lines. (*D*) Association of genes identified in *B* with pathways. EMT, epithelial–mesenchymal transition. (*E*) Integration of N-glycoproteome and transcriptome identifies the ALK-driven cytokine network. Using cytokine receptors with ALK-dependent N-glycosylation as seeds, their protein-to-protein interaction network is constructed. Node color depicts the log$_2$-fold change of transcripts. The gold-colored node border shows the N-glycoproteins with significant depletion after ALK inhibition.

To determine if IL-31Rβ is a functionally relevant target in ALK$^+$ALCL, we silenced IL-31Rβ using lentivirus-mediated RNAi in a DEL cell line. IL-31Rβ silencing (Fig. 3*D*) resulted in decreased clonogenicity of ALK$^+$ALCL cells (Fig. 3*E*). Additionally, tumor growth in a xenograft model was abrogated in ALK$^+$ALCL depleted of IL-31Rβ by RNAi, suggesting that this protein is a functionally relevant therapeutic target in ALK$^+$ALCL (Fig. 3*F*).

CRISPR-Cas9 Screen Reveals Vulnerabilities in Cytokine Receptor Signaling Pathways in ALK$^+$ALCL. Although IL-6–STAT3 and IL-2–STAT5 signaling pathways are identified to be significantly down-regulated by ALK inhibition (Fig. 2*D*), this finding does not establish that they represent therapeutic vulnerabilities in ALK$^+$ALCL. To address this question, we performed a genome-wide functional screen using a lentiviral sgRNA library (9) that targets 14,250 genes in the human genome with an average coverage of three to four sgRNAs per gene in two ALK$^+$ALCL cell lines (SUP-M2 and Karpas 299). Monte Carlo analysis assessing major ALK-dependent signaling pathways in aggregate revealed that the IL-6–STAT3 and IL-2–STAT5 pathways are significant sensitivity targets in ALK$^+$ALCL (Fig. S8*A*). The sgRNA screen showed that members of these pathways, such as OSM (the ligand for IL-31Rβ) and STAT3, are susceptible targets (Fig. S8*B*).

Discussion

The advent of large-scale technologies has dramatically advanced the understanding of cancer pathogenesis. However, most studies involve single-platform profiling approaches that provide a unidimensional view of the pathobiological mechanisms intrinsic to disease states. In contrast, integrative approaches, such as proteogenomics, offer the opportunity to understand better the complex biological networks that underlie the pathogenesis of higher order biological processes, such as cancer. Beyond large-scale annotation of genomic, transcriptomic, and proteomic profiles, the ability to undertake large-scale genetic screens of lethal phenotypes using the CRISPR system greatly facilitates the identification of critical functional genes that could be exploited for precision therapeutics.

We used an MS-based approach to generate a compendium of N-glycoproteins expressed in human lymphomas. Our study revealed a large number of proteins (>1,000) involved in various cellular functions. Importantly, numerous candidate biomarkers and therapeutic targets were identified, and selected candidates were orthogonally and functionally validated.

Unsupervised clustering of the N-glycoproteomic data readily segregated the 32 cell lines based on lineage of origin and respective subtype of lymphoma. Importantly, primary clinical samples were appropriately classified with cell lines representative of their respective cell of origin and lymphoma subtypes based on N-glycoprotein profiles. In all, the N-glycoproteomic signatures were robust and discriminated a wide range of closely related tumor subtypes and yielded several candidate biomarkers for distinct subtypes of lymphoma.

To gain insights into the relationship between RNA expression and N-glycoproteomic profiles and to identify candidate diagnostic
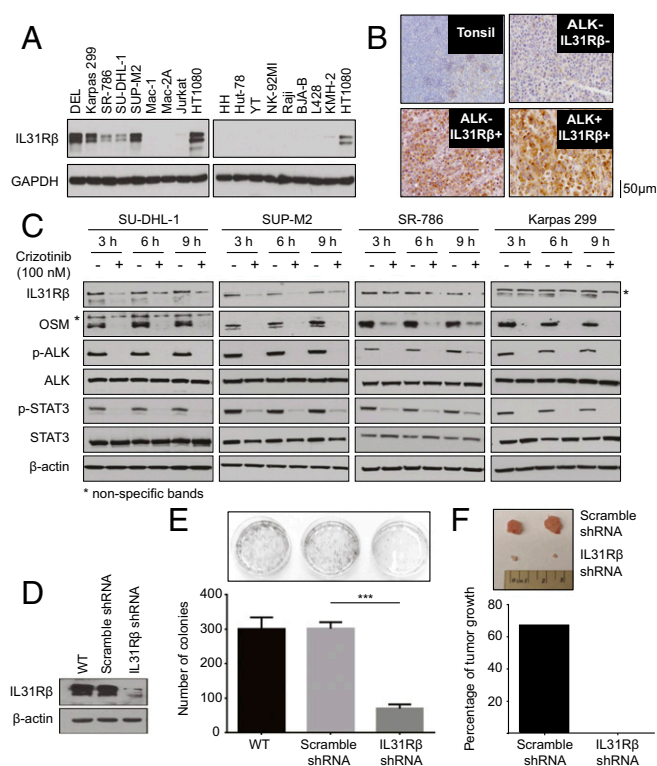
www.manaraa.com

Fig. 3. IL-R31β is selectively expressed in ALK⁺ALCL cells, is regulated by ALK activity, and contributes to NPM-ALK oncogenesis. (*A*) Selective expression of IL-31β in ALK⁺ALCL cell lines was confirmed by Western blotting. (*B*) IL-31β expression in a well-characterized cohort of ALCL patient samples (*n* = 56) demonstrated a statistically significant correlation between ALK and IL-31β expression. (*C*) Western blot results demonstrated that ALK inhibition decreased IL-31Rβ and OSM expression in a time-dependent manner after treatment with crizotinib at 100 nM. The asterisk indicates nonspecific bands. (*D*) Western blot results show IL-31Rβ knockdown in the DEL cell line. (*E*) Stable knockdown of IL-31Rβ in the DEL cell line caused a statistically significant reduction of colony formation after 14 d (mean ± SD, *n* = 3). ***P < 0.001. (*F*) Percentages of tumor growth after mice were xenografted with the modified DEL cell lines.

biomarkers, we focused on ALCL harboring chimeric fusions involving the ALK. We integrated RNA-seq and N-glycoproteomic data to investigate changes mediated by oncogenic ALK activity. This approach highlighted a number of signaling modules regulated by ALK activity, including the inflammatory and IFN-γ and IFN-α responses, as well as pathways regulating hypoxia adaptation. In particular, the top-ranking pathways were the cytokine/receptor regulatory network involving several interleukins, their receptors, and the JAK-STAT signaling axis. Members of this network, including STAT3, were identified by CRISPR susceptibility screens as vulnerability targets in ALK⁺ALCL.

To validate the cytokine/receptor-STAT pathway functionally in ALK⁺ALCL, we followed up on our MS observation of restricted expression of IL-31Rβ in ALK⁺ALCL N-glycoproteomes. The expression of IL-31Rβ, a cytokine 6 receptor family member (24), was confirmed by immunohistochemistry of primary ALK⁺ALCL tissues. Further, we showed that IL-31Rβ is regulated by ALK activity and that IL-31Rβ plays an important role in its pathogenesis and may be used as biomarker for ALK⁺ALCL.

In conclusion, we demonstrated that integration of proteomics and functional genomic analyses yields complementary information that informs our understanding of complex biological processes. Additionally, the N-glycoproteomic data presented herein represent a compendium of candidate biomarkers and a valuable resource for the investigation of the biological roles and functional significance of specific glycoproteins in lymphomas. Taken together, these results suggest a model wherein NPM-ALK–driven signaling promotes an

autocrine–paracrine network through induction of cytokines and their receptors via activation of STAT family transcription factors. Our data suggest that several glycoproteins demonstrate characteristic and unique expression signatures in clinicopathologically distinct forms of human lymphomas. Overall, our results indicate that integration of multidimensional data of cancer cells offers opportunities for novel biomarker discovery and identification of therapeutic targets.

Methods

Cell Lines. All cell lines were grown in RPMI 1644 media supplemented with 10% FBS except for Hut-78, which was grown in Iscove's modified Dulbecco's medium (IMDM) supplemented with 10% FBS.

Clinical Samples. Tumor cells were enriched from peripheral blood of eight patients using an immunomagnetic bead negative selection mixture (EasySep; StellCell Technologies). A total of 671 primary tumor biopsies were selected, and representative areas of different lymphomas were included in tissue microarrays, followed by immunohistochemistry studies. This study was approved by the Institutional Review Board of the University of Michigan (HUM00023256). No informed consent was required for this retrospective study.

N-Linked Glycopeptide Enrichments and LC-MS/MS Analysis. The SPEG protocol previously described was used to isolate N-glycosylated peptides from whole-cell lysates (10). An LTQ OrbitrapXL (ThermoFisher) in-line with Paradigm MS2 HPLC (Michrom BioResources, Inc.) was used for acquiring high-resolution MS and MS/MS data. Each cell line was analyzed in biological replicate format until at least three biological replicates with a glycocapture efficiency of ≥90%, defined as the percentage of identified peptides having at least one deamidated Asn residue, were obtained.

Proteomic Database Searches. RAW files were converted to mzXML using the ReAdW conversion tool from the Trans-Proteomic Pipeline (25) and then searched against the human UniProt database (release 15.15). Searches were carried out using X!Tandem with a *k*-score plug-in and the following parameters: (*i*) precursor parent and daughter ion mass tolerance windows were set to 50 ppm and 0.8 Da, respectively; (*ii*) maximum of two missed cleavages; and (*iii*) variable modifications: oxidized methionine, carbamidomethyl cysteine, and +0.9840 Da on Asn (reflecting conversion of glycosylated Asn to Asp upon PNGaseF-mediated deglycosylation). X!Tandem results were postprocessed using PeptideProphet and ProteinProphet (26, 27).

Extraction of Spectral Counts. Adjusted spectral counts were extracted using Abacus with the following parameters: (*i*) PeptideProphet probability ≥ 0.8, (*ii*) combined file probability ≥ 0.7, and (*iii*) only consider peptides containing modified Asn residue (11). Normalization was performed using a modified version of the normalized spectral abundance factor (NSAF) (28). Our version of NSAF (which we term gNSAF) normalized protein spectral counts to a protein's glycopeptide length. A protein's glyco-length is defined as the length of all its tryptic peptides that contain the glycosylation motif Nx[S/T/C].

Hierarchical Clustering. Throughout the analysis, the R (v3.2.3) (29) package hclust was used for hierarchical clustering and pheatmap was used for generating heat maps. Hierarchical clustering of cell lines was conducted on the log₂-transformed gNSAF values using Euclidean distance and ward.D2 linkage. Given the ordered cell lines, protein expressions were clustered using the Euclidean distance and average linkage. To minimize the influence of low-level signals in the nomination of biomarker candidates, hierarchical clustering was performed with N-glycoproteins with spectral count sums across all cell lines that were greater than the average spectral counts among all the cell lines. A total of 751 N-glycoproteins met this criterion. Pearson correlation between the log₂-transformed gNSAF values of each patient sample and 32 cell lines was calculated and used for clustering based on average linkage of Euclidean distance.

RNA-Seq. Pair-ended RNA-seq reads were aligned to a human reference sequence (February 2009, GRCh37/hg19) using STAR (v2.4.2a) (30). Uniquely mapped reads were quantified by Subread (v1.5.0) and normalized to calculate fragment(s) per kilobase per million reads (FPKM). The log₂-fold change was calculated as the log₂ ratio of CEP-26939 and DMSO-treated samples with a pseudocount of 0.01 after removing the genes with less than 1 FPKM expression. The enrichment of the Hallmark gene sets from MSigDB (v5.1) (20) in the genes with a log₂-fold change lower than −0.8 after ALK inhibition was computed and reported with an FDR q-value cutoff of 0.05. A GSEA preranked tool (v2.2.1) with 10,000 permutations was used with a ranked gene metric of log₂-fold change and the Hallmark gene sets.

Normalization of N-Glycoproteomics Data for ALK-Inhibitor Treatment Experiments. For each cell line, the normalized spectral count of N-glycoproteomics was calculated based on the total spectral count after addition of a pseudocount 1. The normalized spectral count of CEP-26939 versus DMSO treatment was used to calculate the $\log_2$-fold change.

Integration of RNA-Seq and N-Glycoproteomics. A GSEA preranked tool (v2.2.1) with 10,000 permutations was used with a ranked gene metric of $\log_2$-fold change from RNA-seq and the set of genes with $\log_2$-fold change lower than −0.8 in N-glycoproteomics. The $\log_2$-fold changes from RNA-seq and N-glycoproteomics for Karpas 299 and SU-DHL-1 were standardized by the column-wise mean and SD. Hierarchical clustering of genes with Euclidean distance and ward.D2 linkage identified the subgroup of genes simultaneously down-regulated by ALK inhibition in RNA-seq and N-glycoproteomics. The enrichment of the Hallmark gene sets from MSigDB and this subgroup were computed and reported with an FDR cutoff of 0.05. To reconstruct the protein interaction network of cytokine receptors in this set (IL-4R, IL-18R1, IL-1R1, IL-1R2, IL-2RA, and IL-31RB), the STRING database (v10) (31) was used in a Cytoscape (v3.3.0) (32) environment with an interaction confidence cutoff of 0.95.

sgRNA. The result of sgRNA screening was processed with ATARiS (33) to calculate the gene-level vulnerability score in each sample. The minimum of the sgRNA scores in SUP-M2 and Karpas 299 cell lines was assigned to the 14,250 genes included in the screen. For the gene members of the IL-2–STAT5 and the IL-6–STAT3 signaling pathways as defined by MSigDB Hallmark gene sets (34), the sgRNA scores were ranked and plotted alongside the corresponding $\log_2$ FPKM expression level in the DMSO-treated sample. The pathway cumulative vulnerability score of a pathway was defined as the sum of the sgRNA scores of its members. To assess the negative effect of a pathway on cell survivability, Monte Carlo simulation on the population of the sgRNA scores of human genes as defined by the HUGO Gene Nomenclature Committee (35) was performed. In each permutation, a random pathway with the same cardinality of the pathway of interest was constructed and a cumulative pathway vulnerability score was calculated. The distribution of the pathway cumulative vulnerability score of 1E6 permutations was presented and compared with the observed value for the pathway of interest.

Western Blot Analysis. The following primary antibodies were obtained from Cell Signaling Technology: CD44 (clone 156-3C11), anti-ALK (clone C26G7), p-ALK (Y1604; polyclonal), STAT3 (clone 79D7), and p-STAT3 (Y705; clone M9C6). Primary antibodies against CD276 (clone 6A1) and PON2 (clone AF3E6) were purchased from Abcam. Primary antibodies against gp130 (clone M-20) were obtained from Santa Cruz Biotechnology. The antibody against IL-31Rβ was obtained from R&D Systems. The loading quality was assessed using antibodies against GAPDH (clone 6C5; Millipore) or β-actin (clone AC-74; Sigma–Aldrich).

Flow Cytometry Analysis. Flow cytometry analyses were performed using a FACSCanto II (BD Biosciences) flow cytometer, and results were analyzed with FlowJo software. Labeled antibodies were FITC–anti-CD19 (clone SJ25C1; BD Pharmingen), phycoerythrin (PE)–Cy7–anti-CD19 (clone HIB19; BD Pharmingen), PE–Cy7–anti-CD44 (clone G44-26; BD Pharmingen), FITC–anti-CD276 (clone FM276; MiltenyiBiotec GmbH), and APC–anti-IL-31Rβ (clone AN-V2; eBioscience).

IL-31Rβ Knockdown. IL-31Rβ was knocked down in the DEL cell line using the pLKO.1 lentiviral shRNA (TRCN0000289933) from Sigma–Aldrich. A scramble pLKO.1 lentiviral shRNA was used as a control. IL-31Rβ knockdown was assessed by Western blotting.

Colony Formation Assay. DEL cells transformed with scramble shRNA or IL-31Rβ shRNA were incubated for 14 d in methylcellulose-based media (MethoCult; Stem Cell Technology). Colonies were stained with iodonitrotetrazolium chloride overnight and then counted under a microscope. Each assay was performed in triplicate.

Mouse Xenograft. SCID-Beige mice (Charles River Laboratories) were injected s.c. with 10×10^6 cells in the flank (100-μL injection volume containing 50% Matrigel; Becton Dickinson). All animal studies were performed under the compliance of the University of Michigan Committee on the Use and Care of Animals (protocol PRO00003289).

1. Pan S, Chen R, Aebersold R, Brentnall TA (2011) Mass spectrometry based glycoproteomics–from a proteomics perspective. *Mol Cell Proteomics* 10:R110. 003251.
2. Zhang H, Li XJ, Martin DB, Aebersold R (2003) Identification and quantification of N-linked glycoproteins using hydrazide chemistry, stable isotope labeling and mass spectrometry. *Nat Biotechnol* 21:660–666.
3. Roth J (2002) Protein N-glycosylation along the secretory pathway: Relationship to organelle topography and function, protein quality control, and cell interactions. *Chem Rev* 102:285–303.
4. Ramos-Medina R, et al. (2011) Immunohistochemical analysis of HLDA9 Workshop antibodies against cell-surface molecules in reactive and neoplastic lymphoid tissues. *Immunol Lett* 134:150–156.
5. Oflazoglu E, Kissler KM, Sievers EL, Grewal IS, Gerber HP (2008) Combination of the anti-CD30-auristatin-E antibody-drug conjugate (SGN-35) with chemotherapy improves antitumour activity in Hodgkin lymphoma. *Br J Haematol* 142:69–73.
6. Reff ME, et al. (1994) Depletion of B cells in vivo by a chimeric mouse human monoclonal antibody to CD20. *Blood* 83:435–445.
7. Smith MR (2003) Rituximab (monoclonal anti-CD20 antibody): Mechanisms of action and resistance. *Oncogene* 22:7359–7368.
8. Porter DL, Kalos M, Zheng Z, Levine B, June C (2011) Chimeric antigen receptor therapy for B-cell malignancies. *J Cancer* 2:331–332.
9. Wang T, et al. (2015) Identification and characterization of essential genes in the human genome. *Science* 350:1096–1101.
10. Zhou Y, Aebersold R, Zhang H (2007) Isolation of N-linked glycopeptides from plasma. *Anal Chem* 79:5826–5837.
11. Fermin D, Basrur V, Yocum AK, Nesvizhskii AI (2011) Abacus: A computational tool for extracting and pre-processing spectral count data for label-free quantitative proteomic analysis. *Proteomics* 11:1340–1345.
12. Vizcaíno JA, et al. (2013) The PRoteomics IDEntifications (PRIDE) database and associated tools: Status in 2013. *Nucleic Acids Res* 41:D1063–D1069.
13. Hirokawa T, Boon-Chieng S, Mitaku S (1998) SOSUI: Classification and secondary structure prediction system for membrane proteins. *Bioinformatics* 14:378–379.
14. Schwartz D, Gygi SP (2005) An iterative statistical approach to the identification of protein phosphorylation motifs from large-scale data sets. *Nat Biotechnol* 23:1391–1398.
15. Swerdlow SHCE, et al. (2008) *WHO Classification of Tumours of Haematopoietic and Lymphoid Tissues*, ed Vardiman JW (IARC Press, Lyon, France), 4th Ed, Vol 2.
16. Picker LJ, Weiss LM, Medeiros LJ, Wood GS, Warnke RA (1987) Immunophenotypic criteria for the diagnosis of non-Hodgkin's lymphoma. *Am J Pathol* 128:181–201.
17. Fernàndez V, et al. (2010) Genomic and gene expression profiling defines indolent forms of mantle cell lymphoma. *Cancer Res* 70:1408–1418.
18. Shaw AT, et al. (2013) Crizotinib versus chemotherapy in advanced ALK-positive lung cancer. *N Engl J Med* 368:2385–2394.
19. Barrett T, et al. (2013) NCBI GEO: Archive for functional genomics data sets–update. *Nucleic Acids Res* 41:D991–D995.
20. Subramanian A, et al. (2005) Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci USA* 102: 15545–15550.
21. Zhang B, et al.; NCI CPTAC (2014) Proteogenomic characterization of human colon and rectal cancer. *Nature* 513:382–387.
22. Auguste P, et al. (1997) Signaling of type II oncostatin M receptor. *J Biol Chem* 272: 15760–15764.
23. Chiarle R, et al. (2005) Stat3 is required for ALK-mediated lymphomagenesis and provides a possible therapeutic target. *Nat Med* 11:623–629.
24. Tanaka M, Miyajima A (2003) Oncostatin M, a multifunctional cytokine. *Rev Physiol Biochem Pharmacol* 149:39–52.
25. Pedrioli PG (2010) Trans-proteomic pipeline: A pipeline for proteomic analysis. *Methods Mol Biol* 604:213–238.
26. Keller A, Nesvizhskii AI, Kolker E, Aebersold R (2002) Empirical statistical model to estimate the accuracy of peptide identifications made by MS/MS and database search. *Anal Chem* 74:5383–5392.
27. Nesvizhskii AI, Keller A, Kolker E, Aebersold R (2003) A statistical model for identifying proteins by tandem mass spectrometry. *Anal Chem* 75:4646–4658.
28. Paoletti AC, et al. (2006) Quantitative proteomic analysis of distinct mammalian Mediator complexes using normalized spectral abundance factors. *Proc Natl Acad Sci USA* 103:18928–18933.
29. R Core Team (2014) *R: A Language and Environment for Statistical Computing* (F Foundation for Statistical Reporting, Vienna). Available at www.R-project.org/. Accessed February 8, 2016.
30. Dobin A, et al. (2013) STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics* 29:15–21.
31. Szklarczyk D, et al. (2015) STRING v10: Protein-protein interaction networks, integrated over the tree of life. *Nucleic Acids Res* 43:D447–D452.
32. Shannon P, et al. (2003) Cytoscape: A software environment for integrated models of biomolecular interaction networks. *Genome Res* 13:2498–2504.
33. Shao DD, et al. (2013) ATARiS: Computational quantification of gene suppression phenotypes from multisample RNAi screens. *Genome Res* 23:665–678.
34. Liberzon A, et al. (2015) The Molecular Signatures Database (MSigDB) hallmark gene set collection. *Cell Syst* 1:417–425.
35. Eyre TA, et al. (2006) The HUGO Gene Nomenclature Database, 2006 updates. *Nucleic Acids Res* 34:D319–D321.

Rolland et al.

www.manaraa.com