Integrated analysis of glycan and RNA in single cells

Fumi Minoshima[#1], Haruka Ozaki[#2,3], Hiroaki Tateno[1,4*]

[1]Cellular and Molecular Biotechnology Research Institute, National Institute of Advanced Industrial Science and Technology (AIST), Tsukuba Central 6, 1-1-1 Higashi, Tsukuba Ibaraki 305-8566, Japan

[2]Bioinformatics Laboratory, Faculty of Medicine, University of Tsukuba, 1-1-1 Tennodai, Tsukuba, Ibaraki, 305-8575, Japan

[3]Center for Artificial Intelligence Research, University of Tsukuba, 1-1-1 Tennodai, Tsukuba, Ibaraki, 305-8577, Japan

[4]JST PRESTO, Tsukuba Central 6, 1-1-1 Higashi, Tsukuba Ibaraki 305-8566, Japan

[#]These authors contributed equally to this work.

[*]Correspondence should be addressed to H.T. (email: h-tateno@aist.go.jp)

ABSTRACT

Single-cell sequencing has emerged as an indispensable technology to dissect cellular heterogeneity but never been applied to the simultaneous analysis of glycan and RNA. Using oligonucleotide-labeled lectins, we developed a sequencing-based method to jointly analyze glycome and transcriptome in single cells. We analyzed the relationship between the two modalities to uncover the heterogeneity of human induced pluripotent stem cells during differentiation into neural progenitor cells at the single-cell resolution.

Glycans present at the surface of all living cells play crucial roles in diverse biological processes. Glycan structures have been known to vary depending on cell types and states. Therefore, glycans are often referred to as “cell signature” that reflect cellular characteristics. Indeed, most of the stem cell markers[1] and serum tumor markers are glycans[2].

The high structural diversity and complexity of the glycome pose enormous challenges for analytic and functional studies. Glycans are the secondary products of genes synthesized by the orchestration of many proteins, such as glycosyltransferases and glycosidases. There has been no method to predict the precise structure of the complex glycans only from the gene expression profiles. Therefore, it is vital to develop methods to analyze cell surface glycans directly. In this sense, different strategies have been undertaken to analyze glycome[3-5]. Recently, lectin microarray has played a pivotal role in surveying and mapping the informational context of complex glycans of various biological samples, indicating the applicability of lectin-based glycan profiling[6-8]. However, there are limitations to the current glycan analytical methods. For instance, (1) glycans are unable to be analyzed at a single-cell level, (2) the glycan profile of each cell type in the mixed cell populations cannot be obtained without prior cell separation, and (3) the relationship between glycome

and transcriptome in single cells cannot be analyzed. Simultaneous analysis of the two modalities in single cells could lead to the understanding of cellular heterogeneity and glycan functions, and the development of novel glycan markers of rare cells.

High-throughput single-cell sequencing has been transformative to understand the complex cell populations[9]. Recently, simultaneous profiling of multiple types of molecules within a single cell has been developed for building a much more comprehensive molecular view of the cell[9,10]. However, there has been no technology to jointly analyze glycome and transcriptome in single cells since glycans cannot be amplified by PCR, unlike DNA and RNA. Here, we first established highly-multiplexed lectin-based glycan profiling of single cells by sequencing (scGlycan-seq). We then combined the scGlycan-seq with single-cell transcriptome profiling (scRNA-seq) for joint analysis of glycan and RNA in single cells (scGR-seq).

We hypothesized that a lectin conjugated with a DNA oligonucleotide (DNA-barcoded lectin) could be measured by sequencing as a digital readout of glycan abundance. We conjugated lectins to DNA oligonucleotides designed to contain a barcode sequence for the identification of lectin and allowed specific

identification by PCR (Fig. 1a). For efficient immobilization of oligonucleotides with lectins and their subsequent recovery and amplification by PCR, lectins were conjugated with the photocleavable dibenzocyclooctyne-N-hydroxysuccinimidyl ester (DBCO-NHS), which allowed efficient conjugation with 5'-azide-modified oligonucleotides followed by the oligonucleotides to be released from the lectin by ultraviolet (UV) exposure (Supplementary Fig. 1). DNA-barcoded lectins were then purified by affinity chromatography to remove excess DNA oligonucleotides. We prepared a panel of 39 DNA-barcoded lectins that exhibited distinct glycan-binding specificities, and DNA-barcoded mouse and sheep IgG as negative controls (Supplementary Table 1). 41 DNA-barcoded proteins were incubated with $1x10^5$ cells and unbound lectins were removed by washing (Fig. 1b). Then, single-cell or $1x10^4$ cells were separated into PCR tubes and exposed to UV. After centrifugation, supernatants containing released DNA barcodes were recovered, amplified by PCR, and analyzed by next-generation sequencing (NGS) to count the DNA barcodes. Each lectin barcode count was divided by the total lectin barcode count and expressed as a percentage (%) for each lectin.

We first evaluated Glycan-seq using glycoside-polyacrylamide (PAA) probes immobilized on magnetic

beads. When beads were incubated with DNA-barcoded proteins, corresponding lectins showed specific signals, demonstrating specific glycan detection by DNA-barcoded lectins (Supplementary Fig. 2). Next, we assessed the ability of Glycan-seq to discriminate distinct cell populations based on cell surface glycan expression in bulk samples. Obtained data were compared with flow cytometry using fluorescence-labeled lectins as the gold standard. We focused on rBC2LCN, known to specifically bind to human induced pluripotent stem cells (hiPSCs) but not to human dermal fibroblasts (hFibs)[11-13]. A higher percentage of DNA-barcoded rBC2LCN was detected in hiPSCs (42.4%) than hFibs (0.4%), which was consistent with flow cytometry results (Fig. 1c, Supplementary Table 2). Mouse and sheep IgG were used as negative controls, providing negligible DNA barcode levels (<0.01%)(Supplementary Table 2).

Chinese hamster ovary cells and glycosylation-defective mutants (Lec1 and Lec8) were analyzed by Glycan-seq (Supplementary Fig. 3)[14]. Wild-type (WT) cells typically express complex-type N-glycans[15], while Lec8 and Lec1 express agalactosylated and mannosylated N-glycans, respectively. In flow cytometry and Glycan-seq, rLSLN (galactose binder) showed higher binding to WT than Lec8 and Lec1. Similarly, rSRL (GlcNAc binder) and rCalsepa (mannose binder) showed strong binding to Lec8 and Lec1, respectively. These

results demonstrated that Glycan-seq could distinguish different bulk cell populations depending on cell surface glycan expression.

We further addressed whether relative quantitative differences in expression levels observed by flow cytometry could be measured by Glycan-seq. hiPSCs were differentiated into neural progenitor cells (NPCs), which were fluorescently-stained with NPC marker (Nestin and PAX6) antibodies (Supplementary Figs. 4 and 5). In flow cytometry, rBC2LCN binding gradually decreased during differentiation into NPCs (Fig. 1d *left panel*). Similar results were obtained by Glycan-seq (Fig. 1d *right panel*, Supplementary Table2). Therefore, relative differences in glycan profiles observed by Glycan-seq agreed well with flow cytometry.

We then tested scGlycan-seq to see its applicability in single cells, and it revealed heterogeneity of hiPSCs and hFibs (Fig.1e, Supplementary Table3). The hiPSCs after differentiation into NPCs (days 0, 4, and 11) were analyzed by scGlycan-seq. Single-cell heterogeneity of hiPSCs gradually increased after differentiation (Fig. 1f, Supplementary Table3). Moreover, scGlycan-seq recapitulated heterogeneity of rBC2LCN observed by flow cytometry (Supplementary Fig. 6). These demonstrated that scGlycan-seq enabled glycan profiling in

single cells.

The scGlycan-seq was combined with scRNA-seq to enable the simultaneous measurement of glycome and transcriptome (Fig. 1b). We employed RamDA-seq, a full-length single-cell total RNA-sequencing method[7]. We performed scGR-seq on the hiPSCs (n = 53) and hiPSC-derived NPCs (11-day differentiation) (n = 43). Transcriptome data of hiPSCs and hiPSC-derived NPCs revealed cell-type-specific expression of selected cell-type marker genes (Supplementary Fig. 7). Furthermore, we addressed how the information from two molecular layers can be used. When only mRNA data were used, the two cell types (hiPSCs and NPCs) were partially separated in PCA (Fig. 2a *left panel).* This was the case when only glycan data was used for PCA (Fig. 2a *middle panel,* Supplementary Tables 4 and 5). Conversely, in partial least squares (PLS) regression, the two cell types were clearly separated (Fig. 2a *right panel*). This demonstrated how mRNA and glycan information characterize cell identities.

Simultaneous transcriptome and glycome measurements could associate genes with glycans at single-cell level. PLS regression decomposed mRNA and lectin-binding data matrices into different components, and

thereby, mRNAs and lectins were related differently per component (Fig. 2b). For the component p1, where weights were high for rAAL, rBC2LA, rLSLN, and rBanana, the mRNAs (high p1 weight) related to regionalization, sensory organ development, and axonogenesis were the most enriched gene sets (Supplementary Fig. 8), suggesting that the glycan ligands of these lectins might be crossly associated with neural differentiation. This allowed us to infer each glycan's function and role as a marker through the set of genes associated with glycan. Furthermore, by summing the weights of each component, we obtained the relationship between lectins and glycosylation-related genes (Supplementary Fig. 9). For example, *ST6Gal1* showed the highest correlation with rBC2LCN, since *ST6Gal1* products and rBC2LCN ligands were highly expressed in hiPSCs[11]. These exemplify how scGR-seq is useful for finding potential relevance between transcriptome and glycome layers.

The abundance of lectins with similar glycan-binding specificity was highly correlated across single cells (Fig. 2c). We found lectins that fluctuated with each other based on their correlation. For instance, α1-2Fuc-binding lectins, rBC2LCN and TJA2, were associated with the same glycans. This demonstrated that lectin-binding data reflects cell surface glycans. Besides, a pseudo-time analysis was performed utilizing a slingshot

with mRNA data. Separation of single cells indicated that reconstructed pseudo-time reflected cell differentiation progression. Expression of hPSC (*POU5F1*) and NPC (*OTX2*) marker genes showed a progressive decrease and increase along with cell differentiation progression (generalized additive model, $q<0.05$)(Fig. 2d), indicating biological relevance of pseudo-time reconstruction. Next, we searched lectins to unveil which was changed along with the pseudo-time test (Fig. 2e). rBC2LCN was decreased along pseudo-time, while other lectins such as rBanana (mannose binder) were increased. Consistently, rBC2LCN binding was gradually decreased during differentiation into NPCs (Fig. 1d, Supplementary Fig. 10). In contrast, rBanana showed intracellular organella staining in hiPSCs. However, the cell surface was stained for hiPSC-derived NPCs, suggesting the drastic change of the localization of glycans recognized by rBanana during differentiation. Since rBanana showed a high weight in component 1 (Supplementary Fig. 8), the glycan ligands of rBanana might be related to neural differentiation.

We examined genes correlated with hPSC-specific rBC2LCN at the single-cell level (Supplementary Fig. 11 and Supplementary Table 6). The highest positive correlation coefficient was observed with the hPSC marker gene (*POU5F1*), whereas the highest negative was observed with the NPC marker gene (*VIM*),

supporting the previous finding that rBC2LCN was a maker probe for hPSCs[11,12]. In contrast, we investigated lectins correlated with hPSC marker genes (MYC, LEFTY1, and LEFTY2) fluctuating in hiPSCs (Supplementary Fig. 12). rBC2LCN exhibited the highest correlation with MYC as expected. LEFTY1 and LEFTY2 exhibited the highest one with rGal3C (galactose binder) and rBC2LA (mannose binder), respectively, suggesting the possible link between LEFTY1/2 and the glycans ligands of these lectins.

scGR-seq could be readily adaptable to a droplet, micro-well, and indexing-based high-throughput single-cell technologies[9]. Any glycan-binding proteins such as anti-glycan antibodies and glycan-binding peptides could be incorporated in the developed system[16]. We believe that it has the potential to advance the understanding of the cellular heterogeneity as well as the biological roles of glycans across diverse multicellular systems in diverse species.

METHODS

Cells. 201B7 hiPSCs were cultured in 2 ml of mTeSR Plus (VERITAS, Tokyo, Japan) on a 6-well plate coated with Matrigel (Corning International, Tokyo, Japan)[17]. 201B7 hiPSCs were differentiated to neural progenitor

cells using STEMdiff SMADi Neural Induction Kit (VERITAS, Tokyo, Japan).

Conjugation of lectins to DNA oligonucleotides. 100 μg of lectin in 100 μl of PBS was incubated with ten-times the molar amount of dibenzocyclooctyne-N-hydroxysuccinimidyl ester (DBCO-NHS) (Funakoshi Co., Ltd., Tokyo, Japan) at 20°C for 1 h under dark (Supplementary Fig. 1). 10 μl of 1 M Tris was then added and incubated at 20°C for 15 min under dark to inactivate the excess DBCO-NHS. The excess DBCO-NHS was also removed by using Sephadex G-25 desalting columns (GE Healthcare Japan Co., Tokyo, Japan). The resulting DBCO-labeled lectin was then incubated at 4 °C with ten-times the molar amount of 5'-azide-modified DNA oligonucleotides (Integrated DNA Technologies, KK, Tokyo, Japan). To remove unbound oligonucleotides and obtain lectins with sugar-binding activity, the DNA-barcoded lectin was purified by affinity chromatography using appropriate sugar-immobilized Sepharose 4B-CL (GE) based on the glycan-binding specificity of each lectin. Finally, DNA-barcoded lectin was dialyzed by Tube-O-DIALYZER (Takara Bio Inc., Shiga, Japan) and concentrated using centrifugal filters having a 10 kDa molecular weight cut off (MWCO) (Merck KGaA, Darmstadt, Germany). The purified DNA-barcoded lectin was analyzed by SDS-PAGE. Protein and DNA concentration were measured using the Bradford protein assay (Bio-Rad Laboratories, Inc., CA) and Quant-

iT OliGreen ssDNA Reagent (Thermo Fisher Scientific KK, Tokyo, Japan).

List of lectins. See Supplementary Table1 for a list of lectins and barcodes used for the GR-seq.

Glycan-seq in bulk and single cells. Cells ($1x10^5$) were suspended in phosphate-buffered saline (PBS) containing 1% bovine serum albumin (BSA) and incubated with 39 DNA-barcoded lectins at a final concentration of 0.5 μg ml^{-1} at 4°C for 1 h. After washing three times with 1 ml of PBS/BSA, cell number was counted by TC20 auto cell counter (Bio-Rad Laboratories, Inc., CA). Cells ($1x10^4$ cells per tube or 1 cell per tube) were distributed (FCR&BIO Co., LTD., Hyogo, Japan). One cell dispense was performed using TOPick I Live Cell Pick system (Yodaka Giken, Kanagawa, Japan). To liberate oligonucleotides from cells, the cells were irradiated at 365 nm, 15 W for 15 min using UVP Blak-Ray XX-15L UV Bench Lamp (Analytik Jena, Kanagawa, Japan).

The liberated oligonucleotides were then amplified using NEBNext Ultrall Q5 (New England BioLabs Japan Inc., Tokyo, Japan), and i5-index and i7-index primers containing cell barcode sequences (Supplementary Table7). The PCR products were then purified by the Agencour AMPure XP kit (Beckman Coulter, Inc., Tokyo,

Japan), followed by the manufacturer's protocol. The size and the quantity of the PCR products were analyzed by MultiNA (Shimadzu Co., Kyoto, Japan). The PCR products (4 nM from every single cell) were treated with the Miseq Reagent Kit v2 50 Cycles (Illumina KK, Tokyo, Japan) and sequenced by the MiSeq sequencer (26 bp, paired-end) (Illumina KK, Tokyo, Japan).

DNA barcodes derived from lectins were directly extracted from the reads in the FASTQ format. The number of lectin barcodes bound to each cell was counted using the developed software, Barcode DNA counting system (Mizuho Information & Research Institute, Inc., Tokyo, Japan). Each lectin barcode count was divided by the total lectin barcode count and expressed as a percentage (%) for each lectin.

scRNA-seq. Single-cell cDNA library was prepared using the GenNext® RamDA-seqTM Single Cell Kit (Toyobo Co., LTD. Tokyo, Japan), followed by the manufacturer's protocol[18]. The obtained library was then quantified by real-time PCR using the PowerUp SYBR Green Master Mix (Thermo Fisher Scientific KK, Tokyo, Japan) and CFX Connect System (Bio-Rad Laboratories, Inc., Tokyo, Japan). DNA standards for library quantification (Takara Bio Inc., Shiga, Japan) were used as the standard. 10 nM of the cDNA library from every

single cell was treated with the NovaSeq 6000 S4 reagent kit and sequenced by Nova-Seq 6000 (151 bp, paired-end) (Illumina KK, Tokyo, Japan).

Analysis of scRNA-seq data. The FASTQ quality check of reads was performed using FastQC (version 0.11.5). The reads were trimmed by using the fastq-mcf (version 1.0) with the parameters "-L 150 -l 50 -k 4.0 -q 30". The trimmed reads were mapped to the human reference genome (GRCh38, primary assembly) using HISAT2 (version 2.1.0) with the parameters "--no-softclip"[19]. The resulting SAM files were then converted into BAM files using SAMtools (version 1.4). RSeQC (version 1.2) was used for the quality check of the read-mapping[20]. The featureCounts command in Subread (version 1.6.4) was used to quantify gene-level expression with the parameters "-M -O --fraction -p" and the reference gene model (GENCODE v32, primary assembly)[21]. R (version 3.6.1) was used for the statistical analyses and figure generation. The package ropls (version 1.18.8) was used for PLS regression[22].

Analysis of scRNA-seq data. The FASTQ quality check of reads was performed using FastQC (version 0.11.5). The reads were trimmed by using the fastq-mcf (version 1.0) with the parameters "-L 150 -l 50 -k 4.0

-q 30". The trimmed reads were mapped to the human reference genome (GRCh38, primary assembly) using HISAT2 (version 2.1.0) with the parameters "--no-softclip"[19]. The resulting SAM files were then converted into BAM files using SAMtools (version 1.4). RSeQC (version 1.2) was used for the quality check of the read-mapping[20]. The featureCounts command in Subread (version 1.6.4) was used to quantify gene-level expression with the parameters "-M -O --fraction -p" and the reference gene model (GENCODE v32, primary assembly)[21]. R (version 3.6.1) was used for the statistical analyses and figure generation. The package ropls (version 1.18.8) was used for PLS regression[22].

The cells with low mapped reads were discarded from the pseudo-time analysis. Diffusion maps were performed using the destiny package (version 3.0.1) after retaining genes where at least 3 counts were found in at least 10 cells[23]. The pseudo-time analysis was performed using the slingshot package (version 1.4.0)[24]. The package gam (version 1.16.1) was used for fitting a generalized additive model to search for dynamically regulated genes (from highly variable genes (FDR<0.01)) and lectins (q<0.05).

Flow cytometry. Lectins were recombinantly expressed in *Escherichia coli* and purified by affinity

chromatography as described[11]. Lectins were labeled with R-phycoerythrin (PE) using the R-phycoerythrin Labeling Kit (Dojindo Laboratories Co. Ltd., Kumamoto, Japan). $1x10^5$ cells were suspended in 100 µl of PBS/BSA and incubated with 10 µg ml^{-1} of PE-labeled lectins or anti-PAX6 mAb (x20, clone No. O18-1330, BD Biosciences, CA) on ice for 1 h. Flow cytometry data were acquired on a CytoFLEX (Beckman Coulter, Inc., CA) and analyzed using the FlowJo software (FlowJo, LLC., OR).

Fluorescence staining. Cells were fixed with 4% paraformaldehyde at room temperature for 20 min. After blocking with PBS containing 1% BSA and 0.2% Triton-X at room temperature for 30 min, cells were stained with anti-Nestin mAb (x1000, Clone 25/NESTIN (RUO), BD Biosciences, CA) or 1 μg/mL of Cy3-labeled lectins at room temperature for 60-90 min and observed by fluorescence microscope (IX51, Olympus, Tokyo, Japan). Nucleus was stained with Hoechst33342 (x1000, Dojindo Laboratories, Kumamoto, Japan).

Acknowledgments

This work was supported by JST, PRESTO Grant Number JPMJPR16F6, Japan. We thank Tadashi Kimura in National Institute of Advanced Industrial Science and Technology (AIST) for NGS analysis, Haruki Odaka in

AIST for fluorescence staining of neural progenitor cells, Keiko Hiemori, Kayo Kiyoi, and Jinko Murakami in AIST for technical assistance on the single cell analysis. We also thank Mika Yoshimura, Tetsutaro Hayashi, Itoshi Nikaido in Laboratory for Bioinformatics Research, RIKEN Center for Biosystems Dynamics for thoughtful discussion on RamDA-seq data analysis.

Author contributions

HT designed research, analyzed the data, and wrote the paper. FM carried out the experiments and analyzed the data. HO performed data analysis and wrote the paper.

Competing interests

The authors declare no competing interests.

Data availability

All raw data of Glycan-seq is provided in the supplementary table.

All raw data of scRNA-seq has been deposited to the gene expression omnibus (GEO) with the accession code GSE151642.

To review GEO accession GSE151642:

Go

to https://jpn01.safelinks.protection.outlook.com/?url=https%3A%2F%2Fwww.ncbi.nlm.nih.gov%2Fgeo%2Fquery%2Facc.cgi%3Facc%3DGSE151642&data=02%7C01%7Ch-tateno%40aist.go.jp%7Cfbc1091684f94c0fe02c08d807445932%7C18a7fec8652f409b8369272d9ce80620%7C0%7C0%7C637267337437339077&sdata=mClp5hDiDwT5sft9sYkGkPmPQVfj2v4QgTS9GAJCQdg%3D&reserved=0

Enter token srineewsdtandgz into the box

References

1 Tang, C. *et al.* An antibody against SSEA-5 glycan on human pluripotent stem cells enables removal of teratoma-forming cells. *Nat Biotechnol* **29**, 829-834, doi:nbt.1947 [pii]

10.1038/nbt.1947 (2011).

2 Pearce, O. M. T. Cancer glycan epitopes: biosynthesis, structure and function. *Glycobiology* **28**, 670-696 (2018).

3 Haslam, S. M., North, S. J. & Dell, A. Mass spectrometric analysis of N- and O-glycosylation of tissues and cells. *Curr Opin Struc Biol* **16**, 584-591 (2006).

4 Yamaguchi, Y. & Kato, K. Dynamics and Interactions of Glycoconjugates Probed by Stable-Isotope-Assisted Nmr Spectroscopy. *Method Enzymol* **478**, 305-322 (2010).

5 Narimatsu, H. *et al.* Current Technologies for Complex Glycoproteomics and Their Applications to Biology/Disease-Driven Glycoproteomics. *J Proteome Res* **17**, 4097-4112 (2018).

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

7 Hirabayashi, J., Yamada, M., Kuno, A. & Tateno, H. Lectin microarrays: concept, principle and applications. *Chem Soc Rev* **42**, 4443-4458, doi:10.1039/c3cs35419a (2013).

8 Narimatsu, H. *et al.* A strategy for discovery of cancer glyco-biomarkers in serum using newly developed technologies for glycoproteomics. *FEBS J* **277**, 95-105, doi:EJB7430 [pii]

10.1111/j.1742-4658.2009.07430.x (2009).

9 Stuart, T. & Satija, R. Integrative single-cell analysis. *Nat Rev Genet* **20**, 257-272, doi:10.1038/s41576-019-0093-7 (2019).

10 Stoeckius, M. *et al.* Simultaneous epitope and transcriptome measurement in single cells. *Nat Methods* **14**, 865-+ (2017).

11 Tateno, H. *et al.* Glycome diagnosis of human induced pluripotent stem cells using lectin microarray. *J Biol Chem* **286**, 20345-20353, doi:M111.231274 [pii]

10.1074/jbc.M111.231274 (2011).

12 Onuma, Y., Tateno, H., Hirabayashi, J., Ito, Y. & Asashima, M. rBC2LCN, a new probe for live cell imaging of human pluripotent stem cells. *Biochem Biophys Res Commun* **431**, 524-529, doi:S0006-291X(13)00061-2 [pii]

10.1016/j.bbrc.2013.01.025 (2013).

13 Tateno, H. *et al.* Podocalyxin is a glycoprotein ligand of the human pluripotent stem cell-specific probe rBC2LCN. *Stem Cells Transl Med* **2**, 265-273, doi:sctm.2012-0154 [pii]

10.5966/sctm.2012-0154 (2013).

14 Esko, J. D. & Stanley, P. in *Essentials of Glycobiology* (eds rd *et al.*) 627-637 (2015).

15 North, S. J. *et al.* Glycomics profiling of Chinese hamster ovary cell glycosylation mutants reveals N-glycans of a novel size and complexity. *J Biol Chem* **285**, 5759-5775, doi:10.1074/jbc.M109.068353 (2010).

16 McKitrick, T. R. *et al.* Development of smart anti-glycan reagents using immunized lampreys. *Commun Biol* **3** (2020).

17 Takahashi, K. *et al.* Induction of pluripotent stem cells from adult human fibroblasts by defined factors. *Cell* **131**, 861-872, doi:S0092-8674(07)01471-7 [pii]

10.1016/j.cell.2007.11.019 (2007).

18 Hayashi, T. *et al.* Single-cell full-length total RNA sequencing uncovers dynamics of recursive splicing and enhancer RNAs. *Nat Commun* **9**, 619, doi:10.1038/s41467-018-02866-0 (2018).

19 Kim, D., Paggi, J. M., Park, C., Bennett, C. & Salzberg, S. L. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nature Biotechnology* **37**, 907-+ (2019).

20 Wang, L. G., Wang, S. Q. & Li, W. RSeQC: quality control of RNA-seq experiments. *Bioinformatics* **28**, 2184-2185 (2012).

21 Liao, Y., Smyth, G. K. & Shi, W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923-930 (2014).

22 Thevenot, E. A., Roux, A., Xu, Y., Ezan, E. & Junot, C. Analysis of the Human Adult Urinary Metabolome Variations with Age, Body Mass Index, and Gender by Implementing a Comprehensive Workflow for Univariate and OPLS Statistical Analyses.

J Proteome Res **14**, 3322-3335 (2015).

23 Angerer, P. *et al.* destiny: diffusion maps for large-scale single cell data in R. *Bioinformatics* **32**, 1241-1243 (2016).

24 Street, K. *et al.* Slingshot: cell lineage and pseudotime inference for single-cell transcriptomics. *Bmc Genomics* **19** (2018).

Figure Legends

Fig.1 | Glycan-seq enabled glycan profiling in bulk and single cells. (a) Illustration of the DNA-barcoded lectin. **(b)** Schematic representation of scGlycan-seq, scRNA-seq, and scGR-seq. Cells were incubated with DNA-barcoded lectins and separated into single cells. After DNA-barcode was released by UV exposure, the released DNA-barcode in the cell culture supernatants was measured by next-generation sequencing. RNA transcripts were purified from the cell pellet and analyzed by scRNAseq. **(c)** Comparison of sequencing-based glycan profiling and flow cytometry. (*left panel*) Binding of PE-labeled rBC2LCN to hiPSCs (*red line*) and hFibs (*black line*) was analyzed by flow cytometry. Grey: Binding of PE-labeled BSA to hiPSCs (negative control). (*right panel*) Binding of DNA-barcoded rBC2LCN to hiPSCs and hFibs was analyzed by Glycan-seq. The number of DNA-barcode derived

from rBC2LCN was divided by that of the DNA-barcode of all lectins, multiplied by 100, and expressed as percentage (%). Data are shown as average ± SD of triplicates of the same sample. **(d)** (*left panel*) Binding of PE-labeled rBC2LCN to hiPSCs after differentiation to NPCs for 0 (*red line*), 4 (*green line*), and 11 days (*blue line*). (*right panel*) Binding of DNA-barcoded rBC2LCN to hiPSCs after differentiation to NPCs for 0 (*red bar*), 4 (*green bar*), and 11 days (*blue bar*). Data are shown as average ± SD of triplicates of the same sample. **(e)** Principal component analysis (PCA) of bulk (100 cells, n = 3 for each cell type) and single cells (n = 96 for each cell type) of hiPSCs and hFibs. Glycan-seq data are available in Supplementary Tables 2 and 3. **(f)** PCA of bulk (100 cells, n = 3 for each cell type) and single cells (n = 96 for each cell type) of hiPSCs after differentiation to NPCs for 0, 4, and 11 days. DNA-barcode derived from each lectin was divided by DNA-barcode of all lectins and multiplied by 100. Glycan-seq data are available in Supplementary Tables 2 and 3. PC1: principal component 1. PC2: principal component 2.

Fig. 2 | Simultaneous analysis of glycan and RNA by scGR-seq. (a) PCA of hiPSCs (n = 53, *blue*) and hiPSC-derived NPCs (n = 43, *red*) using scRNAseq. scGlycan-seq data are available in

Supplementary Tables 4 and 5. **(b)** Components of PLS regression. **(c)** Correlation between lectins. **(d)** Psuedo-time analysis using a slingshot with mRNA data. Data of the hPSC marker (*POU5F1*) and NPC marker (*OTX2*) genes were shown. **(e)** Psuedo-time analysis using a slingshot with lectin signals. Lectins with $p<0.05$ were shown.

Fig. 1

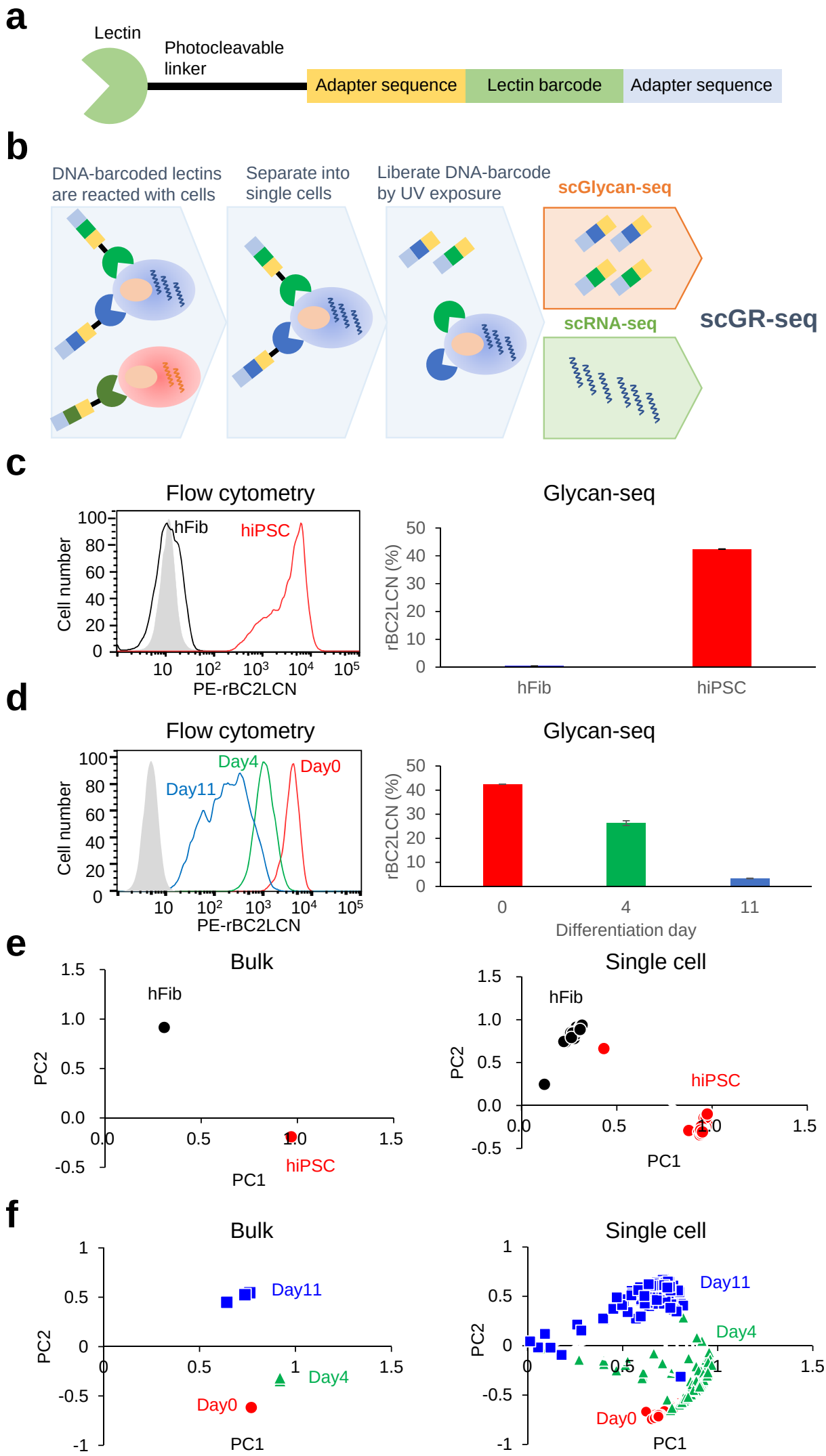
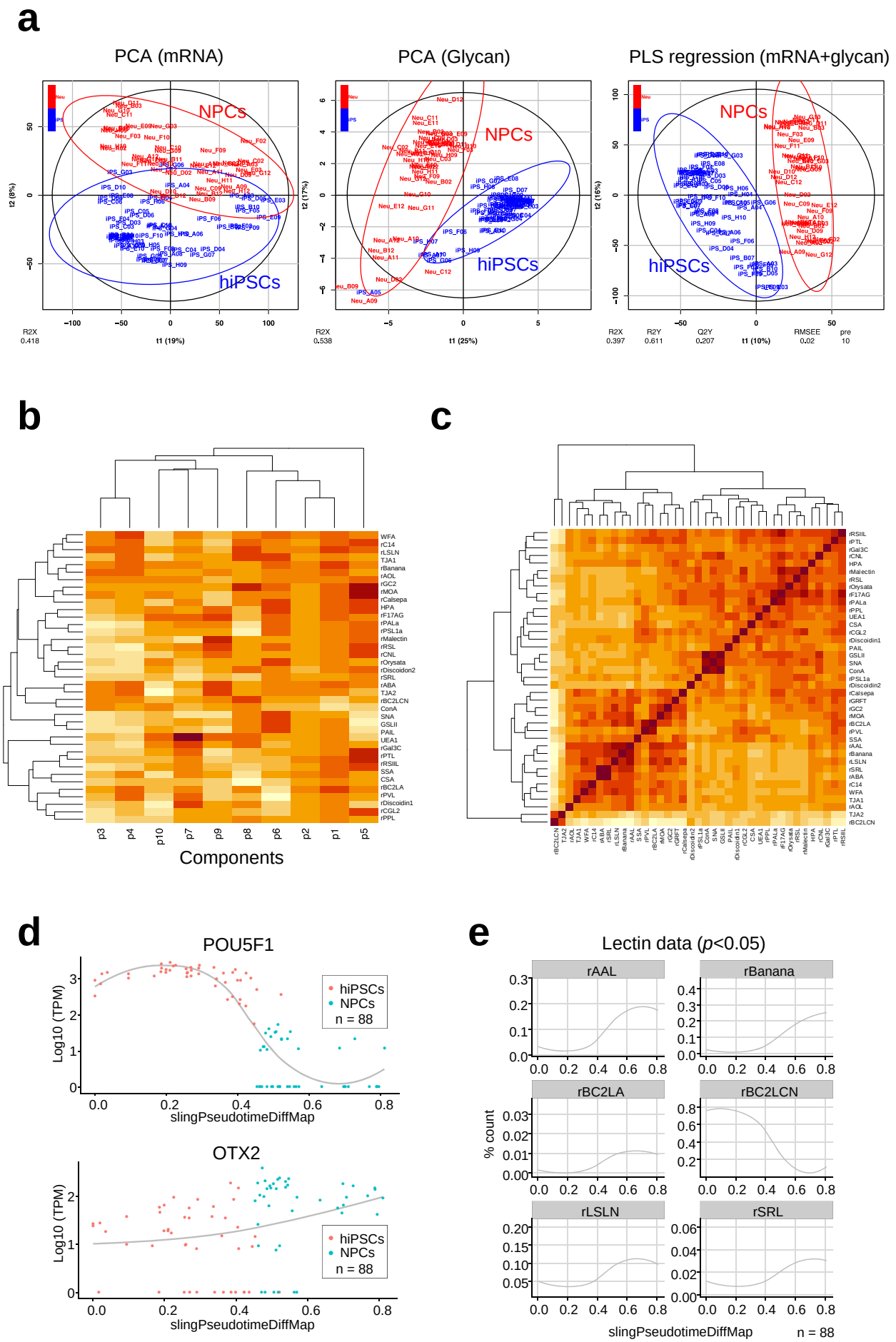

a
Lectin
Photocleavable linker
Adapter sequence
Lectin barcode
Adapter sequence
b
DNA-barcoded lectins are reacted with cells
Separate into single cells
Liberate DNA-barcode by UV exposure
scGlycan-seq
scRNA-seq
scGR-seq
c
Flow cytometry
Cell number
hFib
hiPSC
PE-rBC2LCN
Glycan-seq
rBC2LCN (%)
hFib
hiPSC
d
Flow cytometry
Cell number
Day4
Day0
Day11
PE-rBC2LCN
Glycan-seq
rBC2LCN (%)
Differentiation day
e
Bulk
hFib
hiPSC
PC1
PC2
Single cell
hFib
hiPSC
PC1
PC2
f
Bulk
Day11
Day4
Day0
PC1
PC2
Single cell
Day11
Day4
Day0
PC1
PC2

Fig. 2


a
PCA (mRNA)
PCA (Glycan)
PLS regression (mRNA+glycan)
NPCs
hiPSCs
R2X 0.418
t1 (19%)
t2 (8%)
R2X 0.538
t1 (25%)
t2 (17%)
R2X 0.397
R2Y 0.611
Q2Y 0.207
t1 (10%)
RMSEE 0.02
pre 10
t2 (16%)
b
Components
p3
p4
p10
p7
p9
p8
p6
p2
p1
p5
c
d
POU5F1
OTX2
Log10 (TPM)
slingPseudotimeDiffMap
hiPSCs
NPCs
n = 88
e
Lectin data (p<0.05)
rAAL
rBanana
rBC2LA
rBC2LCN
rLSLN
rSRL
% count
slingPseudotimeDiffMap
n = 88