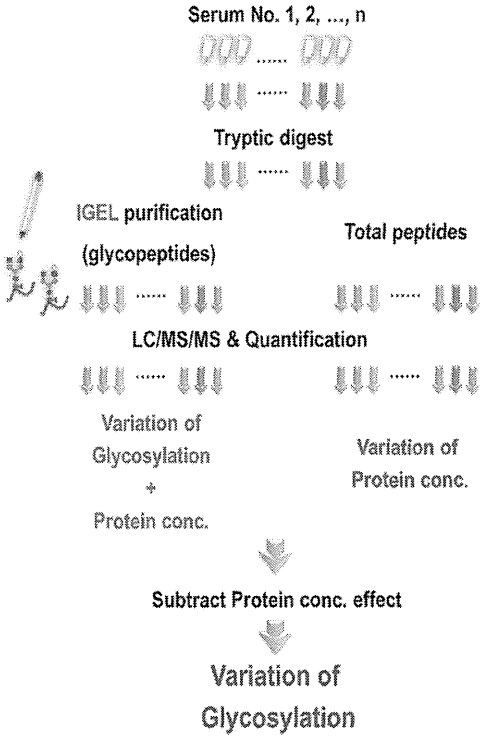

Figure 2. Representative scheme for comprehensive quantification of glycan structure alterations. Following tryptic digestion of serum samples, LC/MS/MS analysis of enriched glycopeptides and total peptides should be performed individually in order to eliminate the effect of core protein concentration variations. Finally, the changing rate of glycans on each peptide could be calculated by subtracting quantification results of total peptides from those of enriched glycopeptides. The representative method for glycopeptide enrichment is IGEL technology illustrated in Fig. 1C.

and pre-enriched nonglycosylated peptides by LC/MS/MS analysis individually. Finally, the site-specific glycoform changes were determined by subtracting core protein concentrations calculated by nonglycosylated peptides from quantification results of each glycopeptide. This approach can illuminate the glycan structure changes on diverse glycosylation sites individually.

2.6 Quantitative glycoproteomic approaches for *O*-glycans

Although tools for comprehensive analysis of *O*-glycosylation are still limited compared with the *N*-glycomics, recent development of sophisticated chemistries have potential to be breakthrough technologies for *O*-glycan biomarker discovery. Hang et al. developed a metabolic labeling approach, which utilized incorporation of tetra-aceylated-*N*-azidoacetylgalactosamine (GalNAz) into the reducing terminus of *O*-glycosylation sites [74, 75]. Following cell culture in the presence of GalNAz for several days, *O*-glycosylated proteins can be specifically collected by alkyne-activated resins (click chemistry) (Fig. 3). Using this technology, Slade et al. identified 267 potentially *O*-glycosylated proteins from the secretome of CHO cells [76]. Furthermore, the Bertozzi's group succeeded to apply GalNAz chemistry to rapid profiling of *O*-linked glycoproteins in living mice [77] and also in vivo imaging of membrane-associated glycans in zebrafish [78]. The other group expanded this technology by using GalNAz, ManNAz, and GlcNAz to discover cell surface differentiation markers on human mesenchymal stem cells [79]. Integration of such metabolic labeling methods for *O*-glycans with mass spectrometric structural analysis may facilitate comprehensive screening of fine *O*-glycan structure alterations in the future.

3 Glycoproteomics for high-throughput biomarker validation

For clinical application of biomarkers, preclinical validation experiments are usually required using independent larger sample set and high-throughput quantification methods. In general, most of protein biomarkers are quantitatively measured by immunoassays in the validation phase, which include sandwich ELISA, AlphaLISA [80–84], and Luminex technology [85–89]. However, it is often extremely difficult to make specific antibodies against both detailed glycan structures and glycosylation sites mainly because of the low immunogenicity of oligosaccharides and structural hindrance of amino acid epitopes by glycans. Thus, alternative glycoproteomic technologies must be established for the replication assays using hundreds of clinical samples, which could detect site-specific glycan structure changes quantitatively with high-throughput manner from complex protein mixtures, such as crude serum/plasma. Here we introduce a couple of technologies used in the glycosylation biomarker validation area.

3.1 Lectin-antibody sandwich ELISA

A sandwich-type ELISA using lectin- and analyte-specific antibody was originally developed by Drouin et al. in 1988 [90]. They intended to establish high-throughput diagnostic assay for Bernard Soulier syndrome, which was a rare autosomal recessive coagulopathy leading to a deficiency of glycoprotein Ib (GpIb), an important clotting regulator as the receptor for von Willebrand factor. In that study they immobilized WGA lectin on 96-well plates to capture GpIb in prepared human platelet proteins and eventually succeeded to detect GpIb by specific monoclonal antibody AN51 quantitatively. They mentioned that this approach was simple, inexpensive, and sensitive way to quantitate glycoproteins for which specific lectins and monoclonal antibodies were available. Tojo et al.

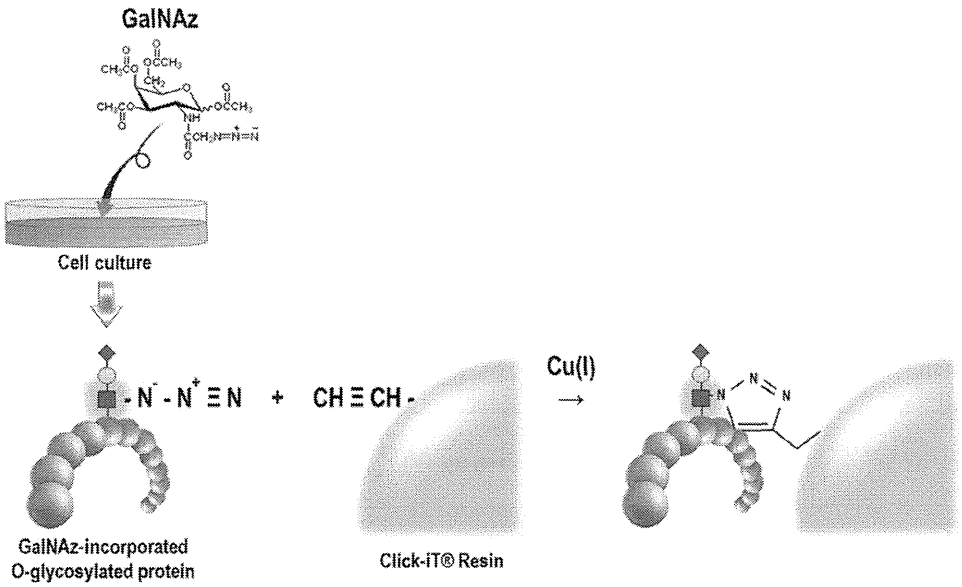

Figure 3. Principle of GalNAz-based *O*-glycosylated protein profiling. Metabolic incorporation of GalNAz residue at the reducing terminus of *O*-glycans enables subsequent capture by alkene reactive resins.

applied the same concept of assay utilizing ConA-immobilized ELISA plates and specific polyclonal antibody to quantitate D-mannans of *Candida albicans* [91]. Since ConA lectin possessed a high binding specificity for the D-mannopyranose unit, the sensitivity and specificity of D-mannan detection were significantly improved compared to previous antibody–antibody ELISA or quantitative precipitin reaction, they reported.

In cancer diagnostic studies, Parker reported the application of lectin/antibody sandwich ELISA assay to the serological diagnosis of pancreatic cancer [92]. His team captured glycoproteins possessing *N*-acetylglucosamie and sialic acid moieties by WGA lectin and detected by CAM17.1 monoclonal antibody specific to a part of mucins. They provided diagnostic assay results from not only retrospective study showing a sensitivity of 78% for pancreatic cancer with a specificity of 76% [93], but also prospective study showing even better sensitivity and specificities (84 and 92%, respectively), suggesting that the assay probably performed better on fresh samples. They also circumstantially summarized the usefulness of CAM 17.1/WGA test for pancreatic cancer diagnosis in *Lancet* journal [94]. Very recently, Miyoshi's group found that fucosylated haptoglobin had a great potential for the detection of pancreatic cancer and prognosis of postoperative colorectal cancer [95, 96]. They coated 96-well plates with antihaptoglobin Fab antibody because IgG had fucosylated oligosaccharides in its Fc portion, and detected fucosylated haptoglobin by biotinylated AAL lectin [95–98]. They measured serum samples from 397 individuals and concluded that the sensitivity and specificity for the diagnosis of pancreatic cancer patients from normal controls was 50 and 91%, respectively [98].

3.2 Energy-resolved oxonium ion monitoring (Erexim) technology

For the purpose of first high-throughput site-specific quantification of glycan structure variations, we recently developed Erexim technology [99]. The oxonium ions are defined as any oxygen cations with three bonds in chemistry, while they are used synonymously with sugar oxonium ions in glycoproteomics, which are produced as oligosaccharide fragment ions in collision cells of mass spectrometers during CID of glycopeptides [100]. Typical oxonium ions (m/z) and the corresponding oligosaccharide components are shown in the right side of Fig. 4. The detection of particular set of oxonium ions in MS/MS spectra of glycopeptides not only represents the existence of glycan modifications but also provides signature of original glycan structures [101]. Furthermore, we found that monitoring the yields of oxonium ions over a wide range of collision energy by use of MRM [102–104] on quadrupole mass spectrometer exhibited a highly glycan structure-unique fragmentation patterns. Indeed, the Erexim curves allowed us to clearly distinguish even structural isomers (same mass with distinct glycan linkage) and separately quantify their contents (Fig. 4). This technology was then applied to the *N*-glycan profiling of three model therapeutic antibody drugs, Herceptin (trastuzumab), Avastin (bevacizumab), and Erbitux (cetuximab). In the lot-to-lot glycan structure variation test for Herceptin or Avastin, around 30 glycan structures on Fc region of antibody drugs were relatively quantified in 30-min analysis, revealing that at most 10% increase or decrease of several glycoforms were observed in both drugs ($n = 4$ for each of four lots). In the case of Avastin glycan profiling, glycans on both Fab and Fc

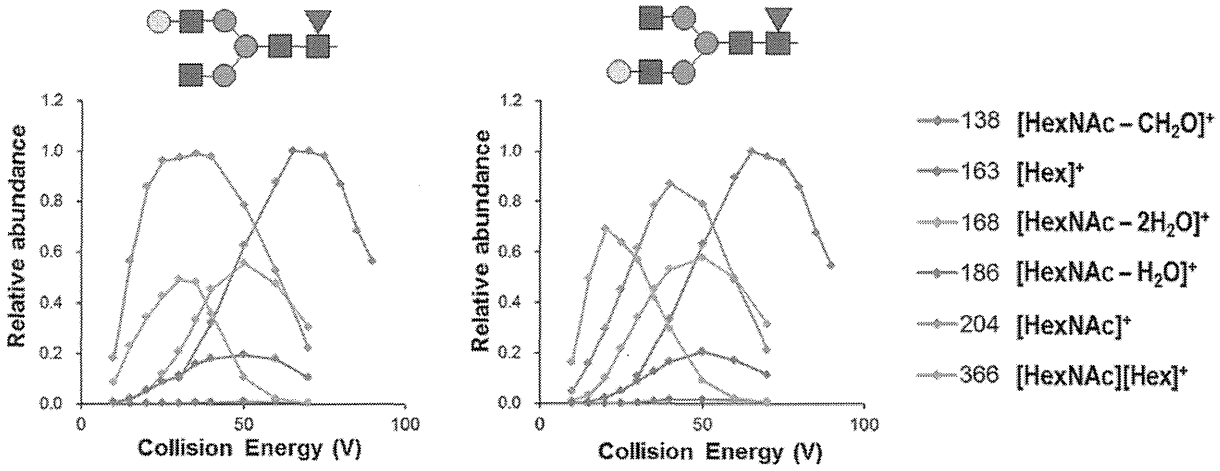

Figure 4. Quantitative evaluation of site-specific glycan structure microheterogeneities by Erexim technology. Representative Erexim curves for a pair of structural isomers detected on IgG molecule are shown. Six oxonium ions listed on the right side were quantitatively monitored by MRM mode. The patterns of collision energy-dependent ion yield curves for $m/z = 204$ and 366 were clearly distinguishable between α1,6-linked galactose (left) and α1,3-linked galactose (right) at the nonreducing termini.

regions were simultaneously quantified in a single Erexim analysis. The result showed that most of glycans on Fab region were nonhuman-type glycan structures possessing *N*-glycolylneuraminic acids (Neu5Gc) or Gal (α1–3) Gal structures, whereas Fc region had conserved nonimmunogenic glycan structures. The LOD and dynamic range of this technology were 30 attomole and more than four orders, respectively. Since the required sample preparation prior to mass spectrometric analysis was only usual trypsin digestion, the Erexim procedure would be appropriate for automated high-throughput analysis. Thus, this technology has enough potential to be applied to routine evaluation of drug quality, safety, and potency, for which extremely high reproducibility, quantitative capability, and throughput. Meanwhile, our Erexim profiling technique is promising technology allowing rapid and site-specific validation of extracted glycan structure-targeted tumor marker candidates using multiple crude specimens.

4 Toward industrialization and approval of carbohydrate-targeting biomarkers

Biomarkers developed for commercial use and regulatory approval must be required to present data supporting validity and clinical utility. According to the FDA Guidance for industry: Pharmacogenomic data submissions 2005, a valid biomarker is a biomarker that is measured in an analytical test system with "well-established performance characteristics" and for which there is an "established scientific framework or body of evidence that elucidates the clinical significance of test results." In case of carbohydrate-targeting biomarkers for cancer early detection or prognosis, the developed devices need to be approved as in vitro diagnostics.

Toward that purpose, a couple of key requirements have to be fulfilled. First, development of high-throughput, easy-to-use, reproducible, and not so expensive diagnostic devices would be required, which are suitable for widespread clinical use. As described in Section 3.1, lectin-antibody sandwich ELISA system could be the first choice satisfying all of these criteria. However, in some cases it would be difficult to achieve sensitive detection for low abundant serum glycoproteins because of insufficient specificity and affinity of lectins for the glycan epitopes. The Erexim technology (Section 3.2) can provide extremely high sensitivity, specificity, and reproducibility for the quantitative assessment of site-specific glycan structure disorders, whereas the high-spec mass spectrometer and proficient skills of LC/MS/MS operation are required.

The other requirement toward the approval of diagnostics is biological evidence explaining how and why the targeted glycan structure alterations would occur on a particular serum glycoprotein in cancer patients. It is often complicated to figure out the origin of biomarker glycoproteins and biochemical mechanisms of abnormal oligosaccharide generation, especially when investigating for glycoproteins identified from serum shotgun proteomics-driven biomarker screening. To present scientifically strong evidences for carbohydrate-targeting biomarker candidates, further basic and detailed glycobiological studies will be necessary, including analysis of cellular glycan biosynthesis pathways, comprehensive expression analysis of glycosyltransferases, or confirmation of glycoform changes in cancer tissues. Future development of more sensitive, high-throughput, and site-specific glycan structure profiling technologies are also fundamental to facilitate clinical application of numerous carbohydrate-targeting biomarker candidates reported previously on papers.

The authors have declared no conflict of interest.

5 References

[1] Anderson, N. L., The clinical plasma proteome: a survey of clinical assays for proteins in plasma and serum. *Clin. Chem.* 2010, *56*, 177–185.

[2] Anderson, N. L., Anderson, N. G., The human plasma proteome: history, character, and diagnostic prospects. *Mol. Cell. Proteomics* 2002, *1*, 845–867.

[3] Ueda, K., Saichi, N., Takami, S., Kang, D. et al., A comprehensive peptidome profiling technology for the identification of early detection biomarkers for lung adenocarcinoma. *PLoS One* 2011, *6*, e18567.

[4] Schulz-Knappe, P., Zucht, H. D., Heine, G., Jurgens, M. et al., Peptidomics: the comprehensive analysis of peptides in complex biological mixtures. *Comb. Chem. High Throughput Screen.* 2001, *4*, 207–217.

[5] Machtejevas, E., Andrecht, S., Lubda, D., Unger, K. K., Monolithic silica columns of various format in automated sample clean-up/multidimensional liquid chromatography/mass spectrometry for peptidomics. *J. Chromatogr. A* 2007, *1144*, 97–101.

[6] Fogli, A., Bulet, P., Peptidomics analysis of lymphoblastoid cell lines. *Methods Mol. Biol.* 2010, *615*, 247–257.

[7] Overall, C. M., Dean, R. A., Degradomics: systems biology of the protease web. Pleiotropic roles of MMPs in cancer. *Cancer Metastasis Rev.* 2006, *25*, 69–75.

[8] Doucet, A., Butler, G. S., Rodriguez, D., Prudova, A. et al., Metadegradomics: toward in vivo quantitative degradomics of proteolytic post-translational modifications of the cancer proteome. *Mol. Cell. Proteomics* 2008, *7*, 1925–1951.

[9] Butler, G. S., Dean, R. A., Smith, D., Overall, C. M., Membrane protease degradomics: proteomic identification and quantification of cell surface protease substrates. *Methods Mol. Biol.* 2009, *528*, 159–176.

[10] Huesgen, P. F., Overall, C. M., N- and C-terminal degradomics: new approaches to reveal biological roles for plant proteases from substrate identification. *Physiol. Plant.* 2012, *145*, 5–17.

[11] Wejda, M., Impens, F., Takahashi, N., Van Damme, P. et al., Degradomics reveals that cleavage specificity profiles of caspase-2 and effector caspases are alike. *J. Biol. Chem.* 2012, *287*, 33983–33995.

[12] Wehr, A. Y., Hwang, W. T., Blair, I. A., Yu, K. H., Relative quantification of serum proteins from pancreatic ductal adenocarcinoma patients by stable isotope dilution liquid chromatography-mass spectrometry. *J. Proteome Res.* 2012, *11*, 1749–1758.

[13] Shi, T., Zhou, J. Y., Gritsenko, M. A., Hossain, M. et al., IgY14 and SuperMix immunoaffinity separations coupled with liquid chromatography-mass spectrometry for human plasma proteomics biomarker discovery. *Methods* 2012, *56*, 246–253.

[14] Koutroukides, T. A., Guest, P. C., Leweke, F. M., Bailey, D. M. et al., Characterization of the human serum depletome by label-free shotgun proteomics. *J. Sep. Sci.* 2011, *34*, 1621–1626.

[15] Beer, L. A., Tang, H. Y., Sriswasdi, S., Barnhart, K. T. et al., Systematic discovery of ectopic pregnancy serum biomarkers using 3-D protein profiling coupled with label-free quantitation. *J. Proteome Res.* 2011, *10*, 1126–1138.

[16] Berven, F. S., Ahmad, R., Clauser, K. R., Carr, S. A., Optimizing performance of glycopeptide capture for plasma proteomics. *J. Proteome Res.* 2010, *9*, 1706–1715.

[17] Jagtap, P., Bandhakavi, S., Higgins, L., McGowan, T. et al., Workflow for analysis of high mass accuracy salivary data set using MaxQuant and ProteinPilot search algorithm. *Proteomics* 2012, *12*, 1726–1730.

[18] Milan, E., Lazzari, C., Anand, S., Floriani, I. et al., SAA1 is over-expressed in plasma of non small cell lung cancer patients with poor outcome after treatment with epidermal growth factor receptor tyrosine-kinase inhibitors. *J. Proteomics* 2012, *76*, 91–101.

[19] Righetti, P. G., Boschetti, E., The ProteoMiner and the FortyNiners: searching for gold nuggets in the proteomic arena. *Mass Spectrom. Rev.* 2008, *27*, 596–608.

[20] Boschetti, E., Righetti, P. G., The ProteoMiner in the proteomic arena: a non-depleting tool for discovering low-abundance species. *J. Proteomics* 2008, *71*, 255–264.

[21] Hartwig, S., Czibere, A., Kotzka, J., Passlack, W. et al., Combinatorial hexapeptide ligand libraries (ProteoMiner): an innovative fractionation tool for differential quantitative clinical proteomics. *Arch. Physiol. Biochem.* 2009, *115*, 155–160.

[22] Fonslow, B. R., Carvalho, P. C., Academia, K., Freeby, S. et al., Improvements in proteomic metrics of low abundance proteins through proteome equalization using ProteoMiner prior to MudPIT. *J. Proteome Res.* 2011, *10*, 3690–3700.

[23] Hekmat, O., He, S., Warren, R. A., Withers, S. G., A mechanism-based ICAT strategy for comparing relative expression and activity levels of glycosidases in biological systems. *J. Proteome Res.* 2008, *7*, 3282–3292.

[24] Kang, X., Sun, L., Guo, K., Shu, H. et al., Serum protein biomarkers screening in HCC patients with liver cirrhosis by ICAT-LC-MS/MS. *J. Cancer Res. Clin. Oncol.* 2010, *136*, 1151–1159.

[25] Ramus, C., Gonzalez de Peredo, A., Dahout, C., Gallagher, M. et al., An optimized strategy for ICAT quantification of membrane proteins. *Mol. Cell. Proteomics* 2006, *5*, 68–78.

[26] Sethuraman, M., McComb, M. E., Huang, H., Huang, S. et al., Isotope-coded affinity tag (ICAT) approach to redox proteomics: identification and quantitation of oxidant-sensitive cysteine thiols in complex protein mixtures. *J. Proteome Res.* 2004, *3*, 1228–1233.

[27] Yan, W., Lee, H., Deutsch, E. W., Lazaro, C. A. et al., A dataset of human liver proteins identified by protein profiling via isotope-coded affinity tag (ICAT) and tandem mass spectrometry. *Mol. Cell. Proteomics* 2004, *3*, 1039–1041.

[28] Li, J., Steen, H., Gygi, S. P., Protein profiling with cleavable isotope-coded affinity tag (cICAT) reagents: the yeast salinity stress response. *Mol. Cell. Proteomics* 2003, *2*, 1198–1204.

[29] Ueda, K., Katagiri, T., Shimada, T., Irie, S. et al., Comparative profiling of serum glycoproteome by sequential

purification of glycoproteins and 2-nitrobenzensulfenyl (NBS) stable isotope labeling: a new approach for the novel biomarker discovery for cancer. *J. Proteome Res.* 2007, *6*, 3475–3483.

[30] Iida, T., Kuyama, H., Watanabe, M., Toda, C. et al., Rapid and efficient MALDI-TOF MS peak detection of 2-nitrobenzenesulfenyl-labeled peptides using the combination of HPLC and an automatic spotting apparatus. *J. Biomol. Tech.* 2006, *17*, 333–341.

[31] Ou, K., Kesuma, D., Ganesan, K., Yu, K. et al., Quantitative profiling of drug-associated proteomic alterations by combined 2-nitrobenzenesulfenyl chloride (NBS) isotope labeling and 2DE/MS identification. *J. Proteome Res.* 2006, *5*, 2194–2206.

[32] Matsuo, E., Toda, C., Watanabe, M., Ojima, N. et al., Selective detection of 2-nitrobenzenesulfenyl-labeled peptides by matrix-assisted laser desorption/ionization-time of flight mass spectrometry using a novel matrix. *Proteomics* 2006, *6*, 2042–2049.

[33] Matsuo, E., Toda, C., Watanabe, M., Iida, T. et al., Improved 2-nitrobenzenesulfenyl method: optimization of the protocol and improved enrichment for labeled peptides. *Rapid Commun. Mass Spectrom.* 2006, *20*, 31–38.

[34] Kuyama, H., Watanabe, M., Toda, C., Ando, E. et al., An approach to quantitative proteome analysis by labeling tryptophan residues. *Rapid Commun. Mass Spectrom.* 2003, *17*, 1642–1650.

[35] Kurogochi, M., Amano, M., Fumoto, M., Takimoto, A. et al., Reverse glycoblotting allows rapid-enrichment glycoproteomics of biopharmaceuticals and disease-related biomarkers. *Angew. Chem.Int. Ed. Engl.* 2007, *46*, 8808–8813.

[36] Kurogochi, M., Matsushista, T., Amano, M., Furukawa, J. et al., Sialic acid-focused quantitative mouse serum glycoproteomics by multiple reaction monitoring assay. *Mol. Cell. Proteomics* 2010, *9*, 2354–2368.

[37] Ziegler, A., Cerciello, F., Bigosch, C., Bausch-Fluck, D. et al., Proteomic surfaceome analysis of mesothelioma. *Lung Cancer* 2012, *75*, 189–196.

[38] Gundry, R. L., Riordon, D. R., Tarasova, Y., Chuppa, S. et al., A cell surfaceome map for immunophenotyping and sorting pluripotent stem cells. *Mol. Cell. Proteomics* 2012, *11*, 303–316.

[39] Danzer, C., Eckhardt, K., Schmidt, A., Fankhauser, N. et al., Comprehensive description of the N-glycoproteome of mouse pancreatic beta-cells and human islets. *J. Proteome Res.* 2012, *11*, 1598–1608.

[40] Bausch-Fluck, D., Hofmann, A., Wollscheid, B., Cell surface capturing technologies for the surfaceome discovery of hepatocytes. *Methods Mol. Biol.* 2012, *909*, 1–16.

[41] Kovarova, H., Gadher, S. J., Wollscheid, B., Focus on stem cell proteomics. *Proteomics* 2011, *11*, 3943–3945.

[42] Wollscheid, B., Bausch-Fluck, D., Henderson, C., O'Brien, R. et al., Mass-spectrometric identification and relative quantification of N-linked cell surface glycoproteins. *Nat. Biotechnol.* 2009, *27*, 378–386.

[43] Liu, L., Zhang, Y., Zhang, L., Yan, G. et al., Highly specific revelation of rat serum glycopeptidome by boronic acid-functionalized mesoporous silica. *Anal. Chim. Acta* 2012, *753*, 64–72.

[44] Xu, Y., Zhang, L., Lu, H., Yang, P., On-plate enrichment of glycopeptides by using boronic acid functionalized gold-coated Si wafer. *Proteomics* 2010, *10*, 1079–1086.

[45] Tang, J., Liu, Y., Qi, D., Yao, G. et al., On-plate-selective enrichment of glycopeptides using boronic acid-modified gold nanoparticles for direct MALDI-QIT-TOF MS analysis. *Proteomics* 2009, *9*, 5046–5055.

[46] Zhu, J., Wang, F., Chen, R., Cheng, K. et al., Centrifugation assisted microreactor enables facile integration of trypsin digestion, hydrophilic interaction chromatography enrichment, and on-column deglycosylation for rapid and sensitive N-glycoproteome analysis. *Anal. Chem.* 2012, *84*, 5146–5153.

[47] Xiong, Z., Zhao, L., Wang, F., Zhu, J. et al., Synthesis of branched PEG brushes hybrid hydrophilic magnetic nanoparticles for the selective enrichment of N-linked glycopeptides. *Chem. Commun.* 2012, *48*, 8138–8140.

[48] Kuo, C. W., Wu, I. L., Hsiao, H. H., Khoo, K. H., Rapid glycopeptide enrichment and N-glycosylation site mapping strategies based on amine-functionalized magnetic nanoparticles. *Anal. Bioanal. Chem.* 2012, *402*, 2765–2776.

[49] Dam, T. K., Gerken, T. A., Brewer, C. F., Thermodynamics of multivalent carbohydrate-lectin cross-linking interactions: importance of entropy in the bind and jump mechanism. *Biochemistry* 2009, *48*, 3822–3827.

[50] Ueda, K., Takami, S., Saichi, N., Daigo, Y. et al., Development of serum glycoproteomic profiling technique; simultaneous identification of glycosylation sites and site-specific quantification of glycan structure changes. *Mol. Cell. Proteomics* 2010, *9*, 1819–1828.

[51] Kaji, H., Saito, H., Yamauchi, Y., Shinkawa, T. et al., Lectin affinity capture, isotope-coded tagging and mass spectrometry to identify N-linked glycoproteins. *Nat. Biotechnol.* 2003, *21*, 667–672.

[52] Kaji, H., Shikanai, T., Sasaki-Sawa, A., Wen, H. et al., Large-scale identification of N-glycosylated proteins of mouse tissues and construction of a glycoprotein database, GlycoProtDB. *J. Proteome Res.* 2012, *11*, 4553–4566.

[53] Liu, T. W., Kaji, H., Togayachi, A., Ito, H. et al., A chemoenzymatic approach toward the identification of fucosylated glycoproteins and mapping of N-glycan sites. *Glycobiology* 2012, *22*, 630–637.

[54] Narimatsu, H., Sawaki, H., Kuno, A., Kaji, H. et al., A strategy for discovery of cancer glyco-biomarkers in serum using newly developed technologies for glycoproteomics. *FEBS J.* 2010, *277*, 95–105.

[55] Sugahara, D., Kaji, H., Sugihara, K., Asano, M. et al., Large-scale identification of target proteins of a glycosyltransferase isozyme by lectin-IGOT-LC/MS, an LC/MS-based glycoproteomic approach. *Sci. Rep.* 2012, *2*, 680.

[56] Han, H., Stapels, M., Ying, W., Yu, Y. et al., Comprehensive characterization of the N-glycosylation status of CD44s by use of multiple mass spectrometry-based techniques. *Anal. Bioanal. Chem.* 2012, *404*, 373–388.

[57] Zhang, S., Liu, X., Kang, X., Sun, C. et al., iTRAQ plus 18O: a new technique for target glycoprotein analysis. *Talanta* 2012, *91*, 122–127.

[58] Kubota, K., Sato, Y., Suzuki, Y., Goto-Inoue, N. et al., Analysis of glycopeptides using lectin affinity chromatography with MALDI-TOF mass spectrometry. *Anal. Chem.* 2008, *80*, 3693–3698.

[59] Snovida, S. I., Bodnar, E. D., Viner, R., Saba, J. et al., A simple cellulose column procedure for selective enrichment of glycopeptides and characterization by nano LC coupled with electron-transfer and high-energy collisional-dissociation tandem mass spectrometry. *Carbohydr. Res.* 2010, *345*, 792–801.

[60] Wada, Y., Tajiri, M., Yoshida, S., Hydrophilic affinity isolation and MALDI multiple-stage tandem mass spectrometry of glycopeptides for glycoproteomics. *Anal. Chem.* 2004, *76*, 6560–6565.

[61] Davies, M., Smith, K. D., Harbin, A. M., Hounsell, E. F., High-performance liquid chromatography of oligosaccharide alditols and glycopeptides on a graphitized carbon column. *J. Chromatogr.* 1992, *609*, 125–131.

[62] Davies, M. J., Smith, K. D., Carruthers, R. A., Chai, W. et al., Use of a porous graphitised carbon column for the high-performance liquid chromatography of oligosaccharides, alditols and glycopeptides with subsequent mass spectrometry analysis. *J. Chromatogr.* 1993, *646*, 317–326.

[63] Fan, J. Q., Kondo, A., Kato, I., Lee, Y. C., High-performance liquid chromatography of glycopeptides and oligosaccharides on graphitized carbon columns. *Anal. Biochem.* 1994, *219*, 224–229.

[64] Thaysen-Andersen, M., Mysling, S., Hojrup, P., Site-specific glycoprofiling of N-linked glycopeptides using MALDI-TOF MS: strong correlation between signal strength and glycoform quantities. *Anal. Chem.* 2009, *81*, 3933–3943.

[65] Lam, M. P., Lau, E., Siu, S. O., Ng, D. C. et al., Online combination of reversed-phase/reversed-phase and porous graphitic carbon liquid chromatography for multicomponent separation of proteomics and glycoproteomics samples. *Electrophoresis* 2011, *32*, 2930–2940.

[66] Selman, M. H., McDonnell, L. A., Palmblad, M., Ruhaak, L. R. et al., Immunoglobulin G glycopeptide profiling by matrix-assisted laser desorption ionization Fourier transform ion cyclotron resonance mass spectrometry. *Anal. Chem.* 2010, *82*, 1073–1081.

[67] Zeng, Z., Hincapie, M., Pitteri, S. J., Hanash, S. et al., A proteomics platform combining depletion, multi-lectin affinity chromatography (M-LAC), and isoelectric focusing to study the breast cancer proteome. *Anal. Chem.* 2011, *83*, 4845–4854.

[68] Yang, Z., Hancock, W. S., Approach to the comprehensive analysis of glycoproteins isolated from human serum using a multi-lectin affinity column. *J. Chromatogr. A* 2004, *1053*, 79–88.

[69] Yang, Z., Hancock, W. S., Monitoring glycosylation pattern changes of glycoproteins using multi-lectin affinity chromatography. *J. Chromatogr. A* 2005, *1070*, 57–64.

[70] Dayarathna, M. K., Hancock, W. S., Hincapie, M., A two step fractionation approach for plasma proteomics using immunodepletion of abundant proteins and multi-lectin affinity chromatography: application to the analysis of obesity, diabetes, and hypertension diseases. *J. Sep. Sci.* 2008, *31*, 1156–1166.

[71] Kullolli, M., Hancock, W. S., Hincapie, M., Preparation of a high-performance multi-lectin affinity chromatography (HP-M-LAC) adsorbent for the analysis of human plasma glycoproteins. *J. Sep. Sci.* 2008, *31*, 2733–2739.

[72] Zheng, X., Wu, S. L., Hincapie, M., Hancock, W. S., Study of the human plasma proteome of rheumatoid arthritis. *J. Chromatogr. A* 2009, *1216*, 3538–3545.

[73] Qiu, R., Regnier, F. E., Use of multidimensional lectin affinity chromatography in differential glycoproteomics. *Anal. Chem.* 2005, *77*, 2802–2809.

[74] Hang, H. C., Yu, C., Kato, D. L., Bertozzi, C. R., A metabolic labeling approach toward proteomic analysis of mucin-type O-linked glycosylation. *Proc. Natl. Acad. Sci. USA* 2003, *100*, 14846–14851.

[75] Laughlin, S. T., Bertozzi, C. R., Metabolic labeling of glycans with azido sugars and subsequent glycan-profiling and visualization via Staudinger ligation. *Nat. Protoc.* 2007, *2*, 2930–2944.

[76] Slade, P. G., Hajivandi, M., Bartel, C. M., Gorfien, S. F., Identifying the CHO secretome using mucin-type O-linked glycosylation and Click-chemistry. *J. Proteome Res.* 2012, *11*, 6175–6186.

[77] Dube, D. H., Prescher, J. A., Quang, C. N., Bertozzi, C. R., Probing mucin-type O-linked glycosylation in living animals. *Proc. Natl. Acad. Sci. USA* 2006, *103*, 4819–4824.

[78] Laughlin, S. T., Baskin, J. M., Amacher, S. L., Bertozzi, C. R., In vivo imaging of membrane-associated glycans in developing zebrafish. *Science* 2008, *320*, 664–667.

[79] Hart, C., Chase, L. G., Hajivandi, M., Agnew, B., Metabolic labeling and click chemistry detection of glycoprotein markers of mesenchymal stem cell differentiation. *Methods Mol. Biol.* 2011, *698*, 459–484.

[80] Foo, J. Y., Wan, Y., Kostner, K., Arivalagan, A. et al., NT-ProBNP levels in saliva and its clinical relevance to heart failure. *PLoS One* 2012, *7*, e48452.

[81] Pfluger, M., Kapuscik, A., Lucas, R., Koppensteiner, A. et al., A combined impedance and AlphaLISA-based approach to identify anti-inflammatory and barrier-protective compounds in human endothelium. *J. Biomol. Screen.* 2012, *18*, 67–74.

[82] Peters, C. D., Jespersen, B., Norregaard, R., AlphaLISA versus ELISA-based detection of interleukin 18 in healthy subjects and patients with end-stage renal disease. *Scand. J. Clin. Lab. Invest.* 2012, *72*, 583–592.

[83] Cosentino, G., AlphaLISA assays to improve the vaccine development process. *Dev. Biol.* 2012, *134*, 107–111.

[84] Waller, H., Chatterji, U., Gallay, P., Parkinson, T. et al., The use of AlphaLISA technology to detect interaction between hepatitis C virus-encoded NS5A and cyclophilin A. *J. Virol. Methods* 2010, *165*, 202–210.

[85] Biancotto, A., Feng, X., Langweiler, M., Young, N. S. et al., Effect of anticoagulants on multiplexed measurement of cytokine/chemokines in healthy subjects. *Cytokine* 2012, *60*, 438–446.

[86] Kim, Y. W., Bae, S. M., Lim, H., Kim, Y. J. et al., Development of multiplexed bead-based immunoassays for the detection of early stage ovarian cancer using a combination of serum biomarkers. *PLoS One* 2012, *7*, e44960.

[87] Rinewalt, D., Shersher, D. D., Daly, S., Fhied, C. et al., Development of a serum biomarker panel predicting recurrence in stage I non-small cell lung cancer patients. *J. Thorac. Cardiovasc. Surg.* 2012, *144*, 1344–1351.

[88] Tait, B. D., Hudson, F., Cantwell, L., Brewin, G. et al., Review article: Luminex technology for HLA antibody detection in organ transplantation. *Nephrology* 2009, *14*, 247–254.

[89] Picascia, A., Infante, T., Napoli, C., Luminex and antibody detection in kidney transplantation. *Clin. Exp. Nephrol.* 2012, *16*, 373–381.

[90] Drouin, J., Izaguirre, C. A., Patenaude, P., Quantitation of cell membrane glycoproteins in pathological conditions using a lectin-bound enzyme-linked immunosorbent assay (ELISA). Application to human platelets in the Bernard-Soulier syndrome. *J. Immunol. Methods* 1988, *110*, 217–223.

[91] Tojo, M., Shibata, N., Osanai, T., Mikami, T. et al., Sandwich enzyme-linked immunosorbent assay of D-mannans of *Candida albicans* NIH A-207 and NIH B-792 strains using concanavalin A and polyclonal rabbit anti-*C. albicans* antisera. *Carbohydr. Res.* 1991, *213*, 325–330.

[92] Parker, N., Lectin/Antibody "sandwich" ELISA for quantification of circulating mucin as a diagnostic test for pancreatic cancer. *Methods Mol. Med.* 1998, *9*, 249–253.

[93] Parker, N., Makin, C. A., Ching, C. K., Eccleston, D. et al., A new enzyme-linked lectin/mucin antibody sandwich assay (CAM 17.1/WGA) assessed in combination with CA 19–9 and peanut lectin binding assay for the diagnosis of pancreatic cancer. *Cancer* 1992, *70*, 1062–1068.

[94] Yiannakou, J. Y., Newland, P., Calder, F., Kingsnorth, A. N. et al., Prospective study of CAM 17.1/WGA mucin assay for serological diagnosis of pancreatic cancer. *Lancet* 1997, *349*, 389–392.

[95] Takeda, Y., Shinzaki, S., Okudo, K., Moriwaki, K. et al., Fucosylated haptoglobin is a novel type of cancer biomarker linked to the prognosis after an operation in colorectal cancer. *Cancer* 2012, *118*, 3036–3043.

[96] Okuyama, N., Ide, Y., Nakano, M., Nakagawa, T. et al., Fucosylated haptoglobin is a novel marker for pancreatic cancer: a detailed analysis of the oligosaccharide structure and a possible mechanism for fucosylation. *Int. J. Cancer* 2006, *118*, 2803–2808.

[97] Miyoshi, E., Shinzaki, S., Moriwaki, K., Matsumoto, H., Identification of fucosylated haptoglobin as a novel tumor marker for pancreatic cancer and its possible application for a clinical diagnostic test. *Methods Enzymol.* 2010, *478*, 153–164.

[98] Matsumoto, H., Shinzaki, S., Narisada, M., Kawamoto, S. et al., Clinical application of a lectin-antibody ELISA to measure fucosylated haptoglobin in sera of patients with pancreatic cancer. *Clin. Chem. Lab. Med.* 2010, *48*, 505–512.

[99] Toyama, A., Nakagawa, H., Matsuda, K., Sato, T. A. et al., Quantitative structural characterization of local N-glycan microheterogeneity in therapeutic antibodies by energy-resolved oxonium ion monitoring. *Anal. Chem.* 2012, *84*, 9655–9662.

[100] Medzihradszky, K. F., Characterization of site-specific N-glycosylation. *Methods Mol. Biol.* 2008, *446*, 293–316.

[101] Scott, N. E., Parker, B. L., Connolly, A. M., Paulech, J. et al., Simultaneous glycan-peptide characterization using hydrophilic interaction chromatography and parallel fragmentation by CID, higher energy collisional dissociation, and electron transfer dissociation MS applied to the N-linked glycoproteome of *Campylobacter jejuni*. *Mol. Cell. Proteomics* 2011, *10*, M000031-MCP000201.

[102] Lemoine, J., Fortin, T., Salvador, A., Jaffuel, A. et al., The current status of clinical proteomics and the use of MRM and MRM(3) for biomarker validation. *Expert Rev. Mol. Diagn.* 2012, *12*, 333–342.

[103] Kitteringham, N. R., Jenkins, R. E., Lane, C. S., Elliott, V. L. et al., Multiple reaction monitoring for quantitative biomarker analysis in proteomics and metabolomics. *J. Chromatogr. B* 2009, *877*, 1229–1239.

[104] Shi, T., Su, D., Liu, T., Tang, K. et al., Advancing the sensitivity of selected reaction monitoring-based targeted quantitative proteomics. *Proteomics* 2012, *12*, 1074–1092.

blood

2013 121: 4340-4347
doi:10.1182/blood-2012-08-446922 originally published online March 28, 2013

Preapoptotic protease calpain-2 is frequently suppressed in adult T-cell leukemia

Makoto Ishihara, Natsumi Araya, Tomoo Sato, Ayako Tatsuguchi, Naomi Saichi, Atae Utsunomiya, Yusuke Nakamura, Hidewaki Nakagawa, Yoshihisa Yamano and Koji Ueda

Blood (print ISSN 0006-4971, online ISSN 1528-0020), is published weekly by the American Society of Hematology, 2021 L St, NW, Suite 900, Washington DC 20036.

Regular Article

LYMPHOID NEOPLASIA

Preapoptotic protease calpain-2 is frequently suppressed in adult T-cell leukemia

Makoto Ishihara,[1] Natsumi Araya,[2] Tomoo Sato,[2] Ayako Tatsuguchi,[1] Naomi Saichi,[1] Atae Utsunomiya,[3] Yusuke Nakamura,[4] Hidewaki Nakagawa,[1] Yoshihisa Yamano,[2] and Koji Ueda[1]

[1]Laboratory for Biomarker Development, Center of Genomic Medicine, RIKEN, Tokyo, Japan; [2]Department of Molecular Medical Science, Institute of Medical Science, St. Marianna University School of Medicine, Kawasaki, Japan; [3]Department of Hematology, Imamura Bun-in Hospital, Kagoshima, Japan; and [4]Section of Hematology/Oncology, Department of Medicine Faculty, The University of Chicago, Chicago, IL

Key Points

- Proteome-wide analysis of HTLV-1–infected T cells identified 17 biomarker proteins for the diagnosis of ATL or HAM/TSP patients.

Adult T-cell leukemia (ATL) is one of the most aggressive hematologic malignancies caused by human T-lymphotropic virus type 1 (HTLV-1) infection. The prognosis of ATL is extremely poor; however, effective strategies for diagnosis and treatment have not been established. To identify novel therapeutic targets and diagnostic markers for ATL, we employed focused proteomic profiling of the $CD4^+CD25^+CCR4^+$ T-cell subpopulation in which HTLV-1–infected cells were enriched. Comprehensive quantification of 14 064 peptides and subsequent 2-step statistical analysis using 29 cases (6 uninfected controls, 5 asymptomatic carriers, 9 HTLV-1–associated myelopathy/tropical spastic paraparesis patients, 9 ATL patients) identified 91 peptide determinants that statistically classified 4 clinical groups with an accuracy rate of 92.2% by cross-validation test. Among the identified 17 classifier proteins, α-II spectrin was drastically accumulated in infected T cells derived from ATL patients, whereas its digestive protease calpain-2 (CAN2) was significantly downregulated. Further cell cycle analysis and cell growth assay revealed that rescue of CAN2 activity by overexpressing constitutively active CAN2 (Δ_{19}CAN2) could induce remarkable cell death on ATL cells accompanied by reduction of α-II spectrin. These results support that proteomic profiling of HTLV-1–infected T cells could provide potential diagnostic biomarkers and an attractive resource of therapeutic targets for ATL. (*Blood.* 2013;121(21):4340-4347)

Introduction

Human T-lymphotropic virus type 1 (HTLV-1) is a human retrovirus that is the pathogenic agent of HTLV-1–associated diseases, such as adult T-cell leukemia (ATL) and HTLV-1–associated myelopathy/tropical spastic paraparesis (HAM/TSP). Recent epidemiological studies revealed that HTLV-1 is endemic mainly in Japan, the Caribbean basin, Iran, Africa, South America, and the Melanesian islands.[1] Other estimates have shown that 20 million to 30 million people worldwide are infected with HTLV-1.[2] The infection is followed by a prolonged asymptomatic phase of 20 to 30 years, and 2% to 5% of the infected individuals develop ATL during their lifetime.[3] ATL is one of the most aggressive hematologic malignancies characterized by increased numbers of lymphocytes with multilobulated nuclei, so-called flower cells, in blood circulation. The prognosis is severe with the median overall survival period and 5-year survival rate of ATL patients of 7 months and 20%, respectively.[4] Recently, humanized anti-CCR4 (KW-0761) therapeutic antibody achieved a great improvement in ATL treatment in a phase 3 study. However, the disease control rate was restricted to 50%, and long-term prognosis has yet to be known.[5] For future improvements in the management of ATL, novel biomarkers for early diagnosis are urgently needed for early therapeutic intervention.

To date, comprehensive genomic or proteomic studies using $CD4^+$ T cells have been performed for this purpose,[6-9] but reproducibility and reliability of quantification results in the discovery phase were uncertain due to the diverse individual variety of HTLV-1–infected cell contents in $CD4^+$ T cells. To overcome the etiologic variety of samples, we focused on the $CD4^+CD25^+CCR4^+$ T-cell subpopulation since Yamano et al[10] recently revealed that HTLV-1 preferentially infected $CD4^+CD25^+CCR4^+$ T cells in both ATL and HAM/TSP patients. By targeting $CD4^+CD25^+CCR4^+$ T cells, we here provide the first quantitative proteome map illustrating molecular disorders in pathogenic human T cells directly associated with the onset or progression of ATL. The comprehensive and comparative interpretation of total proteome in infected cells, especially between asymptomatic HTLV-1 carriers and ATL patients, could immediately lead to specific candidates for biomarkers and drugs.

Another challenge to emphasize in this study is our recently established proteomic profiling technologies. It is indisputable that the greater the number of clinical samples analyzed, the more confidently statistical analysis can be undertaken in order to identify diagnostic markers and druggable targets. Despite this fact, previous proteomics reports could not provide high-throughput quantitative methodologies that were sufficient for dealing with even more than 10 clinical samples, excepting a study utilizing a surface enhanced laser desorption/ionization time of flight mass spectrometer. Although the surface enhanced laser desorption/ionization time of flight method drastically improved the performance in both quantification and throughput, allowing relative quantification

Submitted August 1, 2012; accepted March 25, 2013. Prepublished online as *Blood* First Edition paper, March 28, 2013; DOI 10.1182/blood-2012-08-446922.

The online version of this article contains a data supplement.

The publication costs of this article were defrayed in part by page charge payment. Therefore, and solely to indicate this fact, this article is hereby marked "advertisement" in accordance with 18 USC section 1734.

analysis for 96 samples in several hours, at most only 250 unidentified protein peaks were detectable. In the present study, we integrated the proteomics server for the huge data set "Expressionist" (Genedata A.G., Basel, Switzerland) with high-end mass spectrometers to maximize the quality and quantity of protein catalogs transferred from mass spectrometers. We first describe the discovery phase providing a panel of novel diagnostic molecules from quantification of 14 064 peptides and identification of 4763 proteins. As the functional validation phase, we further examined the physiological potential of an identified diagnostic marker candidate, calpain-2 (CAN2), particularly concerning the association of its activity with survival or progression of ATL cells.

Materials and methods

PBMCs and cell lines

Peripheral blood mononuclear cells (PBMCs) from 6 normal donors, 5 asymptomatic carriers, and 9 HAM/TSP patients used in the screening analysis were collected in the St. Marianna University School of Medicine. Those from 9 ATL patients were collected in the Imamura Bun-in Hospital. PBMCs from 4 ATL patients used for the validation experiments were provided by the Joint Study on Predisposing Factors of ATL Development. The others from 4 HAM/TSP patients were collected in the St. Marianna University School of Medicine. The use of these human specimens in this study was approved by individual institutional ethical committees: the Ethical Committee of Yokohama Institute, RIKEN (approval code Yokohama H22-3); the Ethical Committee of St. Marianna University School of Medicine; the Institutional Review Board of Imamura Bun-in Hospital; and the Ethical Committee of the University of Tokyo (approval code 10-50). This study was conducted in accordance with the Declaration of Helsinki.

SO-4, KOB, and KK1 cells were kindly provided by Dr Yasuaki Yamada, cultured in RPMI 1640 supplemented with 10% fetal bovine serum (Cell Culture Bioscience, Tokyo, Japan), 100 kU/L interleukin 2 (Cell Science & Technology Institute Inc., Tokyo, Japan), and 1 × antibiotic-antimycotic solution (Sigma-Aldrich, MO). Jurkat, SUP-T1, CCRF-CEM, and MOLT-3 cells were cultured in RPMI 1640 supplemented with 10% fetal bovine serum and 1 × antibiotic-antimycotic solution. All cell lines were grown at 37°C in 5% CO_2. $CD3^+CD4^+CD25^+CCR4^+$ T cells were isolated with anti-CD3-FITC (eBioscience, San Diego, CA), anti-CCR4-PE (Becton Dickinson, CA), anti-CD4-Cy7 (eBioscience), and anti-CD25-APC (eBioscience) on a Cell Sorter JSAN (Bay Bioscience, Hyogo, Japan).

Sample preparation for mass spectrometric analysis

The $CD4^+CD25^+CCR4^+$ T cells were washed with phosphate-buffered saline 3 times and lysed in denaturation buffer (8 M urea in 50 mM ammonium bicarbonate). After sonication, reduction with 5 mM tris(2-carboxyethyl) phosphine (Sigma-Aldrich) at 37°C for 30 minutes, and alkylation with 25 mM iodacetamide (Sigma-Aldrich) at room temperature for 45 minutes, lysates were digested with Trypsin GOLD (Promega, WI) with protein/enzyme ratio of 25:1 at 37°C for 12 hours. The digested peptides were desalted with Oasis HLB μElution plate (Waters, MA). The collected samples were dried up with a Vacuum Spin Drier (TAITEC Co. Ltd., Saitama, Japan) and subjected to mass spectrometric analyses.

Liquid chromatography tandem mass spectrometry (LC/MS/MS)

The digested peptides were separated on a 0.1 × 200 mm homemade C_{18} column using a 2-step linear gradient, 2% to 35% acetonitrile for 95 minutes and 35% to 95% acetonitrile for 15 minutes in 0.1% formic acid with a flow rate of 200 nL/min. The eluting peptides were analyzed with a QSTAR-Elite mass spectrometer (AB Sciex, CA) in the smart information-dependent acquisition mode of Analyst QS software 2.0 (AB Sciex). The other parameters on QSTAR-Elite were shown as follows: DP = 60, FP = 265, DP2 = 15, CAD = 5, IRD = 6, IRW = 5, curtain gas = 20, and ion spray voltage = 2000 V.

Two-dimensional (2D) LC/MS/MS

Tryptic digests of $CD4^+CD25^+CCR4^+$ T cells were dissolved in 10 mM ammonium formate in 25% acetonitrile and fractionated by a 0.2 × 250 mm monolith strong cation exchange column (GL Science, Tokyo, Japan). Peptides were eluted with an ammonium formate gradient from 10 mM to 1 M in curve = 3 mode for 70 minutes using a Prominence high-performance liquid chromatography (HPLC) system (Shimadzu Corporation, Kyoto, Japan). The eluate was fractionated into 20 fractions and analyzed individually by LTQ-Orbitrap-Velos mass spectrometer (Thermo Scientific, Bremen, Germany) accompanied with the Ultimate 3000 nano-HPLC system. The fractionated peptide samples were separated with the same gradient used in the QSTAR-Elite system described previously and analyzed by LTQ-Orbitrap-Velos acquiring a full MS scan on Fourier-transition mode with MS resolution = 60 000 and simultaneously MS/MS scans for the 20 most intense precursor ions in each MS spectrum on ion-trap mode with regular resolution. Other important parameters for LTQ-Orbitrap-Velos were as follows: capillary temp = 250, source voltage = 2 kV, MS scan range = mass-to-charge ratio (m/z) 400 to 1600, acquire data dependent CID MS/MS for top-20 intense precursors, and dynamic exclusion enabled during 30 seconds. For protein identification, all MS/MS spectra were searched against SwissProt database version 2012_06 (20 232 human protein sequences) using SEQUEST algorithm on ProteomeDiscoverer 1.3 software (Thermo Scientific) with the following parameters: MS tolerance = 3 ppm, MS/MS tolerance = 0.8 Da, maximum missed cleavages = 2, enzyme = trypsin, taxonomy = *Homo sapiens*, fixed modification = carbamidomethylation on cysteine, and variable modification = oxidation on methionine. We accepted the protein identification satisfying the false discovery rate <1% by Percolator false discovery rate estimation algorithm on ProteomeDiscoverer.

Label-free quantification analysis

The LC/MS/MS raw data were imported into the Expressionist RefinerMS module and subjected to the following data processing and relative quantification steps. The total work flow on the RefinerMS module is shown in supplemental Figure 1 (see the *Blood* Web site). The LC/MS/MS raw data set from 29 clinical samples was displayed in 2D planes (m/z vs retention time [RT]). The chromatogram grid was applied to all planes: scan counts = 10, polynom order = 3, and RT smoothing = 0. The planes were simplified by subtracting background noises using chromatogram chemical noise subtraction: RT window = 50 scans, quantile subtraction = 50%, and RT smoothing = 3 scans. After the noise subtraction, data points with intensity <10 were clipped to zero. The RT variety among 29 planes was adjusted by chromatogram RT alignment: RT transformation window = 0.2 minutes, RT search interval = 5 minutes, m/z window = 0.1 Da, and gap penalty = 1. Peaks were detected by chromatogram summed peak detection: summation window = 5 scans, overlap = 50, minimum peak size = 4 scans, maximum merge distance = 10 points, peak RT splitting = true, intensity profiling = max, gap/peak ratio = 1%, refinement threshold = 5, consistency threshold = 0.8, and signal/noise threshold = 1. The detected peaks were grouped into isotopic clusters derived from each molecule using 2-step chromatogram isotopic peak clustering. The first parameters were as follows: minimum charge = 1, maximum charge = 10, maximum missing peaks = 0, first allowed gap position = 3, RT window = 0.1 minute, m/z tolerance = 0.05 Da, isotope shape tolerance = 10, and minimum cluster size ration = 1.2. The second parameters were as follows: minimum charge = 1, maximum charge = 10, maximum missing peaks = 0, first allowed gap position = 3, RT window = 0.1 minute, m/z tolerance = 0.05 Da, and minimum cluster size ration = 0.6.

Expression vectors and siRNA

For the Δ_{19}CAN2 construct, the *CAPN2* fragment was amplified with primers 5′-CATGTCGACTCCCACGAGAGGGCCATCAAGT-3′ and 5′-CATTCTAGATCAAAGTACTGAGAAACAGAGCC-3′ from pBlueBacIII *CAPN2* and cloned into pEFBOS-Myc. Prior to the overexpression experiments, we confirmed that the sequence of the inserted *CAPN2* fragment was identical to the Mammalian Gene Collection sequence (accession number

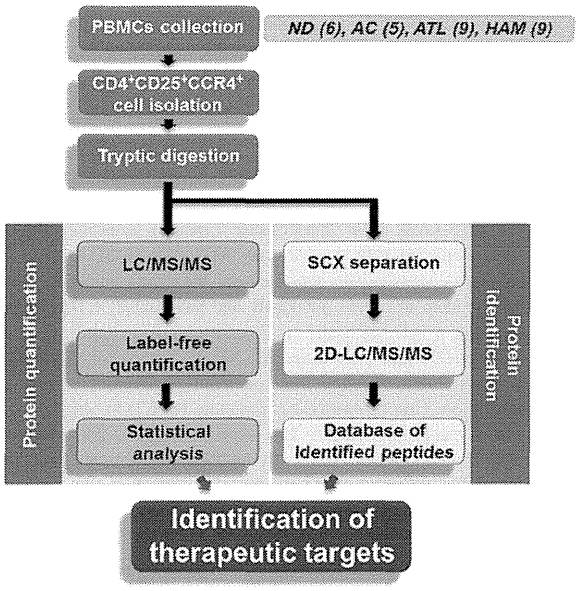

Figure 1. Schematic overview of proteomic profiling for $CD4^+CD25^+CCR4^+$ cells. PBMCs were collected from 6 normal donors, 5 asymptomatic carriers, 9 ATL patients, and 9 HAM/TSP patients, followed by isolation of the $CD4^+CD25^+CCR4^+$ subset using the cell-sorting system. The statistical candidate selection steps, including LC/MS/MS data processing, label-free quantification, and statistical analysis, were performed on the Expressionist proteome server. The protein identification database was separately established based on 2D LC/MS/MS analysis. ND, normal donors; AC, asymptomatic carriers.

BC021303). The 5-μg vector DNA was transfected to 1×10^6 cells. The small interfering RNAs (siRNAs) against *SPTAN1*, *PTMS*, *HSPE1*, and *SHMT2* and siRNA universal negative control were purchased from Sigma-Aldrich. The 500-pmol siRNA oligo was transfected into 1×10^6 cells. The vectors and siRNAs were transfected into all cell lines except CCRF-CEM by Amaxa Nucleoportator transfection Kit V (Lonza, Cologne, Germany) and CCRF-CEM by Kit C (Lonza).

Cell cycle analysis and proliferation assay

For the cell cycle analysis, 1×10^5 to 2×10^5 cells were washed and agitated in 0.1% Triton-X (Sigma-Aldrich) with 100 ng/mL of ribonuclease (Sigma-Aldrich). Following addition of 1 μg/mL propidium iodide, the flow cytometric analysis was performed on FACScalibur (Becton Dickinson). The data analysis was performed using FlowJo software (Tree Star Inc., OR). Doublet events were eliminated from analyses by proper gating on FL2-W/FL2-A primary plots before histogram analysis of DNA content. Cell proliferation was estimated by measuring cell metabolic activity using Cell Counting Kit-8 (Dojindo, Kumamoto, Japan) following the manufacturer's recommendation.

Western blotting

Cells were lysed in lysis buffer [1% NP-40, 2 mM EGTA, 2 mM $MgCl_2$, 150 mM NaCl, 20 mM tris(hydroxymethyl)aminomethane–HCl (pH 7.5), 10% glycerol, containing the protease inhibitor cocktail Complete (Roche, IN)] and subjected to sodium dodecyl sulfate–polyacrylamide gel electrophoresis and transferred onto PVDF membranes. Following blocking with 4% Block Ace (Yukijirushi Nyugyo Inc., Tokyo, Japan), membranes were incubated with anti-myc (9E10; Sigma-Aldrich) or anti–α-II spectrin (Abcam, Cambridge, UK) antibodies. Membranes were then incubated with horseradish peroxidase–conjugated anti-mouse IgG (GE Healthcare, NJ) or anti-rabbit IgG (GE Healthcare), respectively, and visualized with Western Lightning kit (Perkin Elmer, MA).

Multiple reaction monitoring (MRM)

$CD4^+$ T cells were isolated from PBMCs using flow cytometry. The tryptic digests of the isolated cells were analyzed by 4000 Q-TRAP mass spectrometer (AB Sciex) accompanied with Ultimate 3000 nano-HPLC system. The LC gradient was as follows: 2% to 30% acetonitrile for 10 minutes and 30% to 95% acetonitrile for 5 minutes in 0.1% formic acid with a flow rate of 300 nL/min. The MRM transitions monitored were m/z 409.7/375.2 for α-II spectrin (SPTA2); m/z 538.3/889.5 for parathymosin (PTMS); m/z 507.3/147.1 for heat shock 10-kDa protein, mitochondrial (CH10); m/z 490.3/147.1 for serine hydroxymethyltransferase, mitochondrial (GLYM); and m/z 581.3/919.5 for β-actin, respectively. Individual peak areas were normalized by the peak area of β-actin. Data acquisition was performed with ion spray voltage = 2300 V, curtain gas = 10 psi, nebulizer gas = 10 psi, and an interface heating temperature = 150°C. The parameters were set as follows: declustering potential = 60, entrance potential = 10, collision cell exit potential = 10, and dwell time for each transition = 10 seconds. Collision energy was optimized to achieve maximum intensity for each MRM transition as follows: 34.03 V for m/z 409.7/175.1, 24.68 eV for m/z 538.3/889.5, 23.32 eV for m/z 507.3/147.1, 37.57 eV for m/z 490.3/147.1, and 31.58 eV for m/z 581.3/919.5.

Results

Quantitative proteome profiling of $CD4^+CD25^+CCR4^+$ T cells

A schematic overview of the screening approach is shown in Figure 1. To identify diagnostic markers expressed in HTLV-1–infected T cells, a $CD4^+CD25^+CCR4^+$ subset of PBMCs from 6 uninfected volunteers, 5 asymptomatic carriers, 9 HAM/TSP patients, and 9 ATL patients was isolated by flow cytometry (Figure 2). The averaged proportion of $CD4^+CD25^+CCR4^+$ cells in $CD4^+$ T cells from 4 clinical groups was 6.48 ± 2.46%, 13.17 ± 13.06%, 20.55 ± 10.73%, and 55.83 ± 22.40%, respectively, indicating that the occupancy of viral reservoir cells varied drastically among both pathological groups and even individuals within a group. Enrichment of the infected cells was confirmed by viral load measurement of the used samples (supplemental Figure 2). As reported previously,[10] the viral load of $CD4^+CD25^+CCR4^+$ cells (37.91 copies/100 cells on average) was ~10 times higher than that of $CD4^+CD25^-CCR4^-$ cells (4.12 copies/100 cells on average), indicating that the former cells were evidently the HTLV-1–enriched fraction. This fact strongly supports the importance of

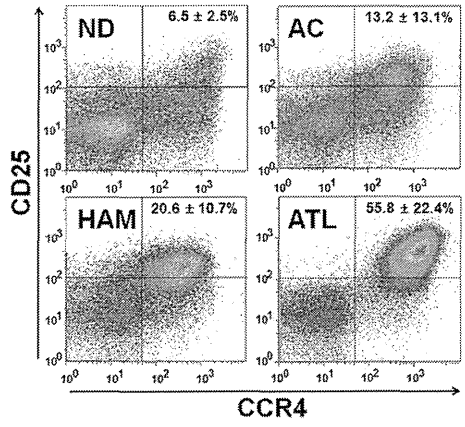

Figure 2. Representative sorting results of $CD4^+CD25^+CCR4^+$ cells. After labeling with anti-CD3-FITC, anti-CD4-Cy7, anti-CD25-APC, and anti-CCR4-PE, the $CD3^+CD4^+CD25^+CCR4^+$ fraction was isolated. The averaged content ± standard deviation (%) of $CD25^+CCR4^+$ cells out of $CD3^+CD4^+$ cells was calculated for each clinical group and is displayed in the upper right section of the panels.

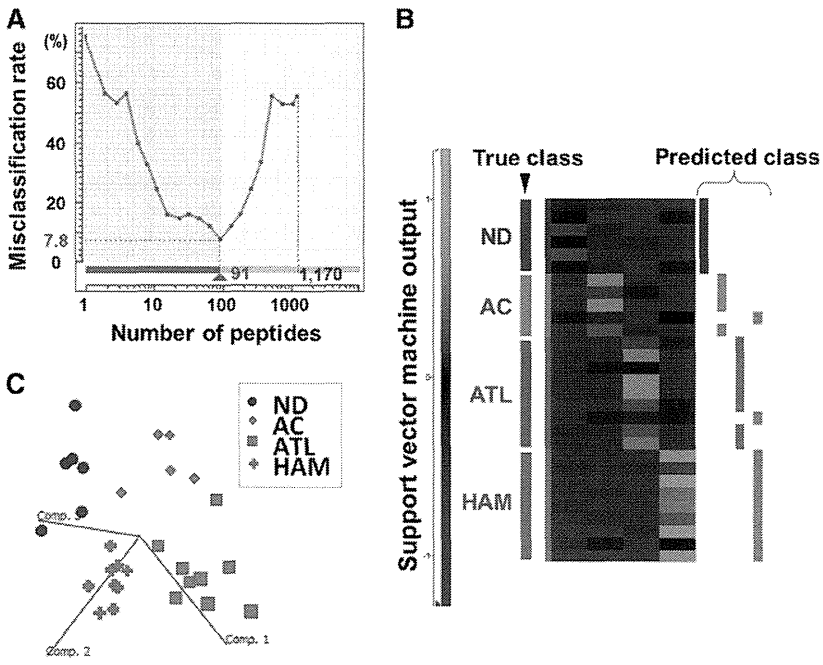

Figure 3. Statistical extraction of candidate therapeutic targets. The 14 064 nonredundant peptides detected were subjected to a 4-group Kruskal-Wallis test (ND, AC, ATL, and HAM), resulting in identification of 1170 first candidates ($P < .01$). ND, normal donors; AC, asymptomatic carriers. (A) Next, the Expressionist ranking method further narrowed down the candidates to 91 peptides based on SVM-REF so that the misclassification rate in the cross-validation test became minimum, 7.8%. (B) The predicted classification result by leave-one-out cross-validation test. The 27 out of 29 cases were successfully classified into the true classes. (C) The three-dimensional plot shows the additional assessment for the classification power of 91 classifiers by principal component analysis. Comp. 1 to 3 indicate principal components 1 to 3.

enriching pathogenic cells for rigorous quantitative biomarker discovery.

An accurately adjusted number of $CD4^+CD25^+CCR4^+$ cells from 29 cases were digested with trypsin and subjected to LC/MS/MS analysis individually. Because recent mass spectrometers often deal with data on the order of hundreds of megabytes per sample, it has been considered almost impossible to calculate a data set larger than a gigabyte from large-scale clinical samples on desktop computers. Hence, we constructed a proteomics server equipped with a 12-core central processing unit, 36 SAS hard disks, and 192-GB physical memories driving the Expressionist, which was designed to combine the database module, the data processing module, and the statistical analysis module into a single integrative platform for genomics, proteomics, and metabolomics. The detailed work flow for data processing and quantification for 29 LC/MS/MS raw data was described in the "Materials and methods" and is illustrated in supplemental Figure 1. Finally, 68 454 nonredundant peaks were detected and grouped into 37 143 isotopic clusters, or molecules. As tryptic peptides should appear as multivalent ions in electrospray ionization mass spectra, 23 079 singly charged ions were removed, resulting in utilization of 14 064 peptide signals for further statistical selection of diagnostic markers.

Statistical identification of candidate diagnostic markers for ATL

A stepwise statistical extraction was employed for the effective identification of proteins, which demonstrated specific up- or downregulation in the ATL group. In the first stage, a 4-group Kruskal-Wallis test was performed to roughly extract the candidates showing a significantly distinct expression level among 4 clinical groups. Here we set the cutoff line at $P < .01$ and obtained 1170 first candidate peptides simply because the isolated peptide set using this criterion showed the best performance in the following prediction model.

Next, we selected the final candidates by the support vector machine–recursive feature elimination algorithm in the Expressionist Analyst module. Support vector machine–recursive feature elimination is a candidate elimination method based on SVM, which enabled us to improve the classification outputs by selecting the best-performing peptide set among initially provided candidates.[11] As a result, a combination of 91 peptides showed the lowest misclassification rate (7.78%) in a leave-one-out cross-validation test (Figure 3A-B). To evaluate the classification efficiency of 91 selected candidates, the principal component analysis was performed. Figure 3C shows the three-dimensional plot of 29 clinical samples based on the 3 best-explainable components, which illustrated statistically clear segregation among the 4 clinical groups. These assessments indicated that the 91 peptides should be a sufficient set of classifiers that closely associated with the pathological characteristics of the 4 clinical groups.

Based on an independently constructed 6279-protein identification database for $CD4^+CD25^+CCR4^+$ cells using 2D LC/MS/MS (see details in "Materials and methods"), 19 peptides among the 91 candidate peptides were successfully assigned to 17 proteins listed in Table 1. The mass spectrometric quantification profiles for the 19 peptides are also shown in Figure 4 (box plots).

Recovering CAN2 activity induced cell death in ATL cells

Our diagnostic marker discovery for ATL identified an enzyme-substrate pair, CAN2 and SPTA2, which demonstrated significantly aberrant expression level in ATL patients (Figure 4). Interestingly, the intensities of the 2 proteins in 27 screening cases (without 2 statistical outliers in Figure 4) showed a clearly inverse correlation ($R^2 = 0.395$, Figure 5A). To examine whether CAN2 downregulation and/or SPTA2 upregulation might be essential for the growth of ATL cells, the enzymatic activity of CAN2 was rescued by overexpressing the constitutively active form of CAN2 (Δ_{19}CAN2) in 3 ATL cell lines, SO-4, KOB, and KK1. After 36 hours of transfection, significant inhibition of cell proliferation (Figure 5B) and induction of sub-G1 transition was observed by activation of CAN2 in 3 ATL cells, but not in 4 non-ATL leukemia cell lines (Figure 5C). Furthermore, overexpression of Δ_{19}CAN2 drastically attenuated the expression level of SPTA2 in the ATL cell

Table 1. List of 17 protein classifiers for categorization of normal donors, asymptomatic carriers, HAM/TSP, and ATL

Accession	Protein name	*P* value (Kruskal-Wallis test)	m/z	RT	Charge	Peptide score	Identity or homology threshold	Sequence
LPPL	Eosinophil lysophospholipase	2.3.E-03	409.722	47.4	2	36.3	27	MVQVWR
CH10	Heat shock 10-kDa protein, mitochondrial	2.5.E-03	430.721	40.6	2	26.2	21	GGIMLPEK
PRG2	Bone marrow proteoglycan	2.4.E-03	528.271	64.6	2	31.6	28	RLPFICSY
MOES	Moesin	8.1.E-04	532.253	26.8	2	46.2	29	EKEELMER
MNDA	Myeloid cell nuclear differentiation antigen	9.4.E-03	647.863	69.1	2	67.3	24	SLLAYDLGLTTK
GLYM	Serine hydroxymethyltransferase, mitochondrial	8.7.E-04	408.551	21.6	3	31.1	18	HADIVTTTTHK
PTMS	Parathymosin	9.7.E-04	453.875	17.8	3	41.2	25	AAEEEDEADPKR
TPIS	Triosephosphate isomerase	9.1.E-03	472.266	71.0	3	54.0	28	QSLGELIGTLNAAK
HSP71	Heat shock 70-kDa protein 1A/1B	9.7.E-03	563.307	65.5	3	93.8	21	IINEPTAAAIAYGLDR
CD6	T-cell differentiation antigen CD6	7.7.E-03	592.306	37.8	3	62.7	22	VLCQSLGCGTAVERPK
ANXA1	Annexin A1	4.4.E-04	612.347	61.5	3	57.0	17	RKGTDVNVFNTILTTR
ANXA6	Annexin A6	2.3.E-03	669.017	70.9	3	54.7	16	AMEGAGTDEKALIEILATR
SPTA2	Spectrin α chain, brain	5.4.E-03	409.718	28.8	2	42.7	30	EAGSVSLR
GLYM	Serine hydroxymethyltransferase, mitochondrial	1.1.E-03	428.240	57.0	2	42.8	27	SGLIFYR
DRB1s	HLA class II histocompatibility antigen, DRB1-1, 4, 10, 11, 13, 15, 16 β chain	1.0.E-02	478.216	25.8	2	55.9	25	AAVDTYCR
CAN2	Calpain-2 catalytic subunit	2.4.E-03	483.253	54.0	2	66.6	29	SDTFINLR
STAT1	Signal transducer and activator of transcription 1-α/β	7.3.E-03	486.290	21.7	2	39.1	29	KILENAQR
PRG2	Bone marrow proteoglycan	9.4.E-04	497.742	49.2	2	31.6	27	FQWVDGSR
CXCL7	Platelet basic protein	1.3.E-03	528.761	43.1	2	51.7	28	ICLDPDAPR

line SO-4 (Figure 5D), but not in the non-ATL leukemia cell line Jurkat (Figure 5E). On the other hand, an additional cell proliferation assay using siRNA against *SPTAN1* revealed that reduction of SPTA2 was not sufficient for the induction of cell death for ATL cells (supplemental Figures 3 and 4).

In addition, 3 proteins (PTMS, CH10, and GLYM) were also found to be upregulated in ATL cells. To address the roles of these proteins, a cell proliferation assay was conducted using 3 ATL cell lines treated with siRNAs against *PTMS*, *HSPE1* (gene symbol of CH10), or *SHMT2* (gene symbol of GLYM) (supplemental Figure 4). As a result, suppression of the *SHMT2* gene induced significant growth inhibition for all 3 ATL cell lines. Although si*HSPE1*-treated KOB cells showed a statistically significant decrease in cell growth rate, si*HSPE1* and si*PTMS* had only partial

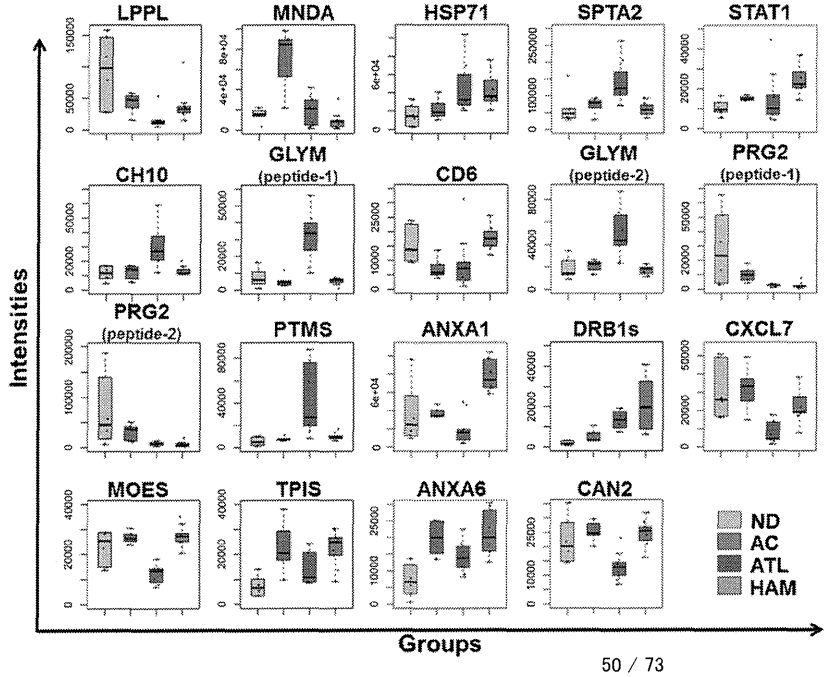

Figure 4. Summary of quantitative features for the 17 protein classifiers identified. The 19 box plots (see Table 1 for protein names) show the results of mass spectrometric quantification and protein identification. We finally identified 19 peptides out of 91 candidates in Figure 3, which were assigned to 17 proteins. Proteins identified from 2 distinct peptides were shown as GLYM (peptides 1 and 2) or PRG2 (peptides 1 and 2). The y-axis indicates normalized relative intensity of peptides in mass spectrometric data. ND, normal donors; AC, asymptomatic carriers.

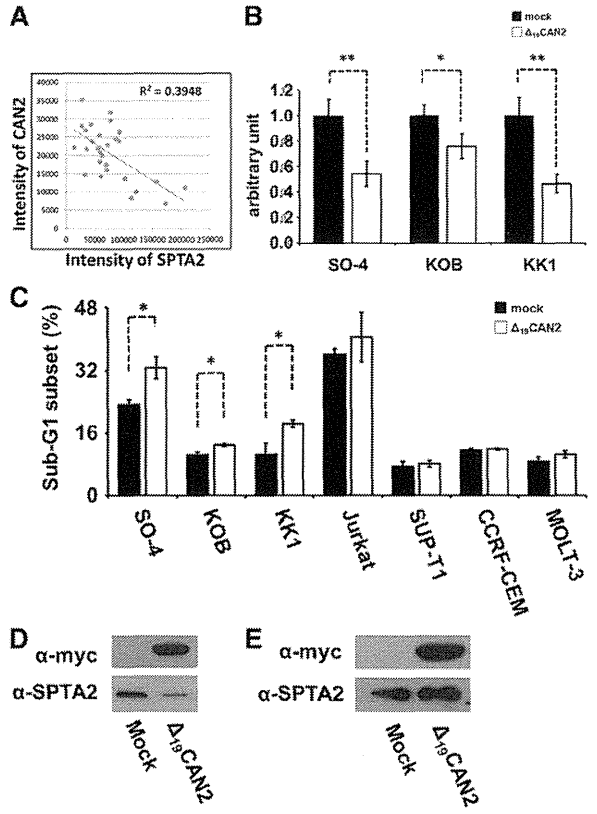

Figure 5. Rescue of CAN2 activity induced cell death in ATL cells. (A) Correlation between CAN2 and SPTA2 expression level in 27 cases. (B) Cell proliferation was measured by MTT assay on SO-4, KOB, and KK1 cells 36 hours after transfection of mock vector or Δ_{19}CAN2. $^{*}P < .05$; $^{**}P < .01$ by Student *t* test. (C) Overexpression of Δ_{19}CAN2 significantly accelerated cell death in 3 ATL (SO-4, KOB, and KK1) and 4 non-ATL (Jurkat, SUP-T1, CCRF-CEM, and MOLT-3) cell lines. $^{**}P < .05$ by Student *t* test. The drastic attenuation of SPTA2 expression was observed after transfection of Δ_{19}CAN2 in SO-4 cells (D), but not in Jurkat cells (E). The immunoblot of anti-myc tag confirmed the expression of exogenous Δ_{19}CAN2.

or no effects on proliferation of ATL cell lines. To further confirm whether the overexpression of SPTA2, PTMS, CH10, or GLYM protein would be an ATL-specific molecular signature, the expression levels of these proteins in 8 clinical samples were evaluated by the mass spectrometric quantification technology MRM (supplemental Figures 5 and 6). Expression of SPTA2, GLYM, and CH10 in cells derived from ATL patients was significantly higher than that in cells derived from HAM/TSP patients. The level of PTMS also showed a clearly increasing tendency in the ATL patient group. Taken together, these results suggested that the deprivation of CAN2 activity and upregulation of GLYM in HTLV-1–infected T cells might have a key role at the onset or progression of ATL.

Discussion

In the past decade, proteomics technologies have developed dramatically for the purpose of obtaining more and more comprehensive and sensitive proteome maps in cells or clinical specimens. The performance of mass spectrometers in particular has exhibited remarkable progress; however, as for sensitivity and throughput, it has still been difficult to identify biomarkers from crude samples including body fluids or total cell lysate. A major reason could be that the range of protein concentration in the analyte is indeed much larger than the dynamic range of recent mass spectrometers.[12] The other essential factor to be improved for clinical proteomics is the capacity of the bioinformatics platform to allow analysis of a sufficient number of clinical samples in order to statistically overcome the significant individual variability.[13]

Concerning the first issue, we previously developed and applied various focused proteomic applications targeting molecular biochemical features including glycan structure biomarkers[14-16] and low-molecular-weight peptide biomarkers.[17] The preenrichment of subproteome fractions effectively reduces the complexity of crude samples and allowed us to identify potential serum cancer biomarkers successfully. Through our previous knowledge, we provide an approach for investigating infectious diseases by employing virus-infected cell-focused proteomics. In addition to HTLV-1, for instance, isolation of HIV-infected cells is highly desired because the frequency of these cells in AIDS patients' PBMCs is ~1 out of 10^4 to 10^5 cells.[18] Actually, we successfully demonstrated the effect of HTLV-1–infected cell isolation on the elimination of individual variability (Figure 2, supplemental Figure 2) and reliable identification of disease state–associated proteins (Figures 4 and 5). We further showed the potential of the next-generation bioinformatics platform Expressionist to remove the constraint on the capacity of data size acquired from high-end mass spectrometers. Expressionist covered whole discovery steps from processing of raw mass spectrometer data to statistical analyses (Figures 1 and 3, and supplemental Figure 1) and, importantly, could perform quantification analysis using a basically unlimited number of clinical samples. Hence, in parallel with the development of mass spectrometers, high-specification and inexpensive OMICS server systems are necessary for future diagnostic marker and therapeutic target discoveries using hundreds or thousands of clinical specimens.

In this study, we focused on the $CD4^+CD25^+CCR4^+$ T-cell subpopulation in which T helper 2, T helper 17, and regulatory T (Treg) cells were mainly involved.[10] The purpose for which we used this subset was to technically enrich the preferential viral reservoir cells and to strengthen reliability of screening results. However, investigating proteome behaviors of these subtypes in HTLV-1–associated diseases is also important physiologically because it has been frequently reported that deregulated Treg plays significant roles in pathogenesis of ATL and HAM/TSP. Indeed, aberrant proliferation of Treg cells is considered the main cause of immunodeficiency in ATL patients because of their innate immunosuppressive functions,[19] whereas abnormal production of interferon γ from infected Treg cells might induce chronic spinal inflammation in HAM/TSP patients.[20] Given the list of our 17 classifier proteins, activation of signal transducer and activator of transcription 1-α/β is the well-known key factor for HAM/TSP,[21] whereas upregulation of heat shock 70-kDa protein 1A/1B, CH10, and PTMS were reported in many other types of tumors.[22-24] The association of these 4 proteins with the etiology of HAM/TSP and ATL would be evident according to the previous work, supporting that our other candidates might similarly have a direct impact on the transformation of Treg cells after infection of HTLV-1. Particularly, the specific upregulation of GLYM in ATL cells represents the first evidence that excessive folate metabolism might be essential for the progression or survival of ATL cells because GLYM is a fundamental enzyme catalyzing the supply of glycine accompanying the conversion of tetrahydrofolate to 5,10-methylenetetrahydrofolate.[25] Indeed, the suppression of GLYM expression, which was confirmed to be upregulated in ATL patients, resulted in significant reduction of cell growth. This observation suggests that diminishing GLYM

expression or enzyme activity could be a promising strategy for molecular-targeting treatment of ATL. Together with the downregulation of CAN2 in the ATL cells shown in Figure 5, the proteins listed in Table 1 could provide the molecular basis for not only interpretation of physiological mechanisms in ATL or HAM/TSP but also development of novel therapeutic agents for HTLV-1–associated diseases.

CAN2 belongs to a Ca^{2+}-regulated cytosolic cysteine protease family, which includes 14 calpain isoforms.[26] The enzymatic activity of calpain is implicated in diverse physiological processes, such as cytoskeletal remodeling, cellular signaling, and apoptosis.[26] As an example of a spectrin-mediated apoptosis pathway, it was reported that CAN2 produced SPTA2 breakdown products following traumatic brain injury.[27] Because SPTA2 interacts with calmodulin and constructs the membrane cytoskeletons, its breakdown is considered a process of membrane structural changes during cell death.[28,29] This fact is concordant with our finding in ATL, suggesting that accumulation of SPTA2 in ATL cells can be attributed to the suppression of CAN2 expression and contribute to circumvent apoptosis. In the analysis of basal levels of CAN2 and SPTA2 in 7 cell lines (supplemental Figure 7), 3 ATL cell lines showed endogenous expression of CAN2 and moderate levels of SPTA2. On the other hand, 4 non-ATL leukemia cells demonstrated very high expression of SPTA2 and undetectable levels of CAN2. Although we found the downregulation of CAN2 and accumulation of SPTA2 in ATL cells, this tendency might be more distinctive in HTLV-1 (–) leukemia cells. Taken together, even though the expression level of CAN2 was indeed suppressed in ATL cells, the CAN2-SPTA2 apoptotic pathway itself might remain normal. In contrast, this pathway was considered to be impaired at multiple stages in HTLV-1 (–) leukemia cells because CAN2 expression was completely diminished (supplemental Figure 7) and overexpression of CAN2 could not reactivate the CAN2-SPTA2 apoptotic pathway (Figure 5B-E). In these cells, not only genetic downregulation of CAN2 but also inhibition of CAN2 enzymatic activity might be involved in the carcinogenesis.

In conclusion, comprehensive proteomic profiling of HTLV-1–infected T cells provided 17 disease-associated signature proteins, which have great potential for future clinical use as diagnostic biomarkers. As we described regarding the relationship between the CAN2-SPTA2 pathway and ATL phenotypes, further individual functional analyses will contribute to understanding the detailed molecular mechanisms involved in the onset or progression of HAM/TSP and ATL.

Acknowledgments

The authors thank Dr Hiroyuki Sorimachi for kindly providing pBlueBacIIICAPN2 vector.

This work was supported by Research on Measures for Intractable Diseases, the Ministry of Health Labour and Welfare Japan.

Authorship

Contribution: M.I. and K.U. designed the study, performed experiments, analyzed results, and wrote the manuscript; A.T. and N.S. performed experiments; N.A., T.S., A.U., and Y.Y. collected the clinical samples and performed flow cytometric experiments; Y.N. and H.N. revised the manuscript; and all authors discussed the results and commented on the manuscript.

Conflict-of-interest disclosure: The authors declare no competing financial interests.

Correspondence: Koji Ueda, Laboratory for Biomarker Development, Center for Genomic Medicine, RIKEN, General Research Building 6F, Institute of Medical Science, 4-6-1 Shirokanedai, Minato-ku, Tokyo, Japan, 1088639; e-mail: k-ueda@riken.jp.

References

1. Yamashita M, Ido E, Miura T, Hayami M. Molecular epidemiology of HTLV-I in the world. *J Acquir Immune Defic Syndr Hum Retrovirol.* 1996;13(suppl 1):S124-S131.
2. Asquith B, Zhang Y, Mosley AJ, et al. In vivo T lymphocyte dynamics in humans and the impact of human T-lymphotropic virus 1 infection. *Proc Natl Acad Sci U S A.* 2007;104(19):8035-8040.
3. Sakashita A, Hattori T, Miller CW, et al. Mutations of the p53 gene in adult T-cell leukemia. *Blood.* 1992;79(2):477-480.
4. Beltran B, Quiñones P, Morales D, Cotrina E, Castillo JJ. Different prognostic factors for survival in acute and lymphomatous adult T-cell leukemia/lymphoma. *Leuk Res.* 2011;35(3):334-339.
5. Ishida T, Joh T, Uike N, et al. Defucosylated anti-CCR4 monoclonal antibody (KW-0761) for relapsed adult T-cell leukemia-lymphoma: a multicenter phase II study. *J Clin Oncol.* 2012; 30(8):837-842.
6. Semmes OJ, Cazares LH, Ward MD, et al. Discrete serum protein signatures discriminate between human retrovirus-associated hematologic and neurologic disease. *Leukemia.* 2005;19(7):1229-1238.
7. Kirk PD, Witkover A, Courtney A, et al. Plasma proteome analysis in HTLV-1-associated myelopathy/tropical spastic paraparesis. *Retrovirology.* 2011;8:81.
8. Rahman S, Quann K, Pandya D, Singh S, Khan ZK, Jain P. HTLV-1 Tax mediated downregulation of miRNAs associated with chromatin remodeling factors in T cells with stably integrated viral promoter. *PLoS ONE.* 2012;7(4):e34490.
9. Polakowski N, Gregory H, Mesnard JM, Lemasson I. Expression of a protein involved in bone resorption, Dkk1, is activated by HTLV-1 bZIP factor through its activation domain. *Retrovirology.* 2010;7:61.
10. Yamano Y, Araya N, Sato T, et al. Abnormally high levels of virus-infected IFN-gamma+ CCR4+ CD4+ CD25+ T cells in a retrovirus-associated neuroinflammatory disorder. *PLoS ONE.* 2009;4(8):e6517.
11. Oh JH, Gao J, Nandi A, Gurnani P, Knowles L, Schorge J. Diagnosis of early relapse in ovarian cancer using serum proteomic profiling. *Genome Inform.* 2005;16(2):195-204.
12. Anderson NL, Anderson NG. The human plasma proteome: history, character, and diagnostic prospects. *Mol Cell Proteomics.* 2002;1(11): 845-867.
13. Nordon IM, Brar R, Hinchliffe RJ, Cockerill G, Thompson MM. Proteomics and pitfalls in the search for potential biomarkers of abdominal aortic aneurysms. *Vascular.* 2010;18(5):264-268.
14. Ueda K, Katagiri T, Shimada T, et al. Comparative profiling of serum glycoproteome by sequential purification of glycoproteins and 2-nitrobenzensulfenyl (NBS) stable isotope labeling: a new approach for the novel biomarker discovery for cancer. *J Proteome Res.* 2007;6(9): 3475-3483.
15. Ueda K, Fukase Y, Katagiri T, et al. Targeted serum glycoproteomics for the discovery of lung cancer-associated glycosylation disorders using lectin-coupled ProteinChip arrays. *Proteomics.* 2009;9(8):2182-2192.
16. Ueda K, Takami S, Saichi N, et al. Development of serum glycoproteomic profiling technique; simultaneous identification of glycosylation sites and site-specific quantification of glycan structure changes. *Mol Cell Proteomics.* 2010;9(9): 1819-1828.
17. Ueda K, Saichi N, Takami S, et al. A comprehensive peptidome profiling technology for the identification of early detection biomarkers for lung adenocarcinoma. *PLoS ONE.* 2011;6(4): e18567.
18. Bahbouhi B, al-Harthi L. Enriching for HIV-infected cells using anti-gp41 antibodies indirectly conjugated to magnetic microbeads. *Biotechniques.* 2004;36(1): 139-147.
19. Matsubar Y, Hori T, Morita R, Sakaguchi S, Uchiyama T. Delineation of immunoregulatory properties of adult T-cell leukemia cells. *Int J Hematol.* 2006;84(1):63-69.
20. Best I, López G, Verdonck K, et al. IFN-gamma production in response to Tax 161-233, and frequency of CD4+ Foxp3+ and Lin HLA-

DRhigh CD123+ cells, discriminate HAM/TSP patients from asymptomatic HTLV-1-carriers in a Peruvian population. *Immunology.* 2009; 128(1, pt 2):e777-e786.

21. Nakamura N, Fujii M, Tsukahara T, et al. Human T-cell leukemia virus type 1 Tax protein induces the expression of STAT1 and STAT5 genes in T-cells. *Oncogene.* 1999;18(17): 2667-2675.
22. Alaiya AA, Al-Mohanna M, Aslam M, et al. Proteomics-based signature for human benign prostate hyperplasia and prostate adenocarcinoma. *Int J Oncol.* 2011;38(4): 1047-1057.
23. Cappello F, Rappa F, David S, Anzalone R, Zummo G. Immunohistochemical evaluation of PCNA, p53, HSP60, HSP10 and MUC-2 presence and expression in prostate carcinogenesis. *Anticancer Res.* 2003;23(2B): 1325-1331.
24. Letsas KP, Vartholomatos G, Tsepi C, Tsatsoulis A, Frangou-Lazaridis M. Fine-needle aspiration biopsy-RT-PCR expression analysis of prothymosin alpha and parathymosin in thyroid: novel proliferation markers? *Neoplasma.* 2007; 54(1):57-62.
25. Anderson DD, Quintero CM, Stover PJ. Identification of a de novo thymidylate biosynthesis pathway in mammalian mitochondria. *Proc Natl Acad Sci USA.* 2011;108(37): 15163-15168.
26. Storr SJ, Carragher NO, Frame MC, Parr T, Martin SG. The calpain system and cancer. *Nat Rev Cancer.* 2011;11(5):364-374.
27. Liu MC, Akle V, Zheng W, et al. Comparing calpain- and caspase-3-mediated degradation patterns in traumatic brain injury by differential proteome analysis. *Biochem J.* 2006;394(pt 3): 715-725.
28. Wallis CJ, Wenegieme EF, Babitch JA. Characterization of calcium binding to brain spectrin. *J Biol Chem.* 1992;267(7): 4333-4337.
29. Liu X, Van Vleet T, Schnellmann RG. The role of calpain in oncotic cell death. *Annu Rev Pharmacol Toxicol.* 2004;44: 349-370.

NEOPLASIA
www.neoplasia.com

Volume 15 Number 11 November 2013 pp. 1251–1261 **1251**

The Oncogenic Polycomb Histone Methyltransferase EZH2 Methylates Lysine 120 on Histone H2B and Competes Ubiquitination[1,2]

Masaharu Kogure*,†, Masashi Takawa*, Vassiliki Saloura‡, Kenbun Sone‡, Lianhua Piao‡, Koji Ueda§, Reem Ibrahim*, Tatsuhiko Tsunoda¶, Masanori Sugiyama†, Yutaka Atomi†, Yusuke Nakamura‡ and Ryuji Hamamoto*,‡

*Laboratory of Molecular Medicine, Human Genome Center, Institute of Medical Science, The University of Tokyo, Minato-ku, Tokyo, Japan; †Department of Surgery, Kyorin University School of Medicine, Mitaka, Tokyo, Japan; ‡Section of Hematology/Oncology, Department of Medicine, The University of Chicago, Chicago, IL; §Laboratory for Biomarker Development, RIKEN, Minato-ku, Tokyo, Japan; ¶Laboratory for Medical Informatics, RIKEN, Yokohama, Kanagawa, Japan

Abstract

The histone methyltransferase enhancer of zeste 2 (EZH2) is known to be a polycomb protein homologous to *Drosophila* enhancer of zeste and catalyzes the addition of methyl groups to histone H3 at lysine 27 (H3K27). We previously reported that EZH2 was overexpressed in various types of cancer and plays a crucial role in the cell cycle regulation of cancer cells. In the present study, we demonstrated that EZH2 has the function to monomethylate lysine 120 on histone H2B (H2BK120). EZH2-dependent H2BK120 methylation in cancer cells was confirmed with an H2BK120 methylation-specific antibody. Overexpression of EZH2 significantly attenuated the ubiquitination of H2BK120, a key posttranslational modification of histones for transcriptional regulation. Concordantly, knockdown of EZH2 increased the ubiquitination level of H2BK120, suggesting that the methylation of H2BK120 by EZH2 may competitively inhibit the ubiquitination of H2BK120. Subsequent chromatin immunoprecipitation–Seq and microarray analyses identified downstream candidate genes regulated by EZH2 through the methylation of H2BK120. This is the first report to describe a novel substrate of EZH2, H2BK120, unveiling a new aspect of EZH2 functions in human carcinogenesis.

Neoplasia (2013) 15, 1251–1261

Introduction

Enhancer of zeste 2 (EZH2) belongs to polycomb group (PcG) protein and is a member of the polycomb repressor complex 2 (PRC2) that methylates histone H3 at lysine 27 (H3K27), a repressive mark that maintains epigenetic silencing of genes. EZH2 is active only when it is associated with other PRC2 core components embryonic ectoderm development (EED), suppressor of zesta 12 homolog (SUZ12), and RBBP4 (retinoblastoma binding protein 4; RbAp48). The PRC2 complex is responsible for repressing of a large number of genes that are essential for development and differentiation. PcG proteins specify positional information such as antero-posterior patterning, through activating or repressing the stable state of *Hox* gene expression. In addition to these well-established functions in embryonic developmen

Address all correspondence to: Ryuji Hamamoto, PhD, Section of Hematology/Oncology, Department of Medicine, The University of Chicago, 5835 S Cottage Grove Ave, MC2115 Chicago, IL 60637. E-mail: rhamamoto@medicine.bsd.uchicago.edu

[1]This work was supported by a grant-in-aid for Young Scientists (22681030) from the Japan Society for the Promotion of Science and Project for Development of Innovation Research on Cancer Therapeutics.

[2]This article refers to supplementary materials, which are designated by Tables W1 to W6 and Figures W1 to W8 and are available online at www.neoplasia.com.

Received 6 August 2013; Revised 16 October 2013; Accepted 21 October 2013

DOI 10.1593/neo.131436

a series of studies have suggested that PcG proteins may influence both Hox-dependent and Hox-independent downstream pathways that control cell proliferation. We previously demonstrated that EZH2 was overexpressed in various types of human cancer, and its overexpression was correlated with a negative outcome in patients with non-small-cell lung carcinoma (NSCLC) after surgical resection [1]. Importantly, selective inhibitors targeting EZH2 have recently developed, and they showed the growth-inhibitory effects of tumor cells [2,3]. These results indicate that EZH2 is a promising target for development of cancer treatment. Although the transcriptional regulation mechanism by EZH2 through methylation of H3K27 has already been well studied, additional functions of EZH2 through methylation of other substrates still remained unclear.

Posttranslational modification of the four core histones is a commonly important process during the regulation of gene activation and repression. Histone modifications are also involved in various cellular processes, including DNA damage response and alternative splicing. Histone H2B is one of the four core histones involved in chromatin formation in eukaryotic cells. Featuring a main globular domain and a long N-terminal tail, H2B is involved with the structure of the nucleosomes of the "beads on a string" structure. As posttranslational modifications of histone H2B, acetylation, phosphorylation, ubiquitination, and sumoylation have already been reported [4–8]. Among these modifications, histone H2B lysine 120 monoubiquitination (H2BK120ub) is known to be a key histone modification that plays critical roles in the transcriptional regulation as well as higher order chromatin organization in many species [9]. H2BK120ub is associated with a high level of gene expression in human cells [10]. This histone modification is also induced after DNA damage and has been indicated to have a critical role in the maintenance of replication-dependent histone mRNA 3′-end processing [11]. The human ring finger protein 20 (RNF20)/RNF40 complex is the major H2B E3 ligase [12]. At the structure level, monoubiquitination of H2BK120 interferes with compaction of chromatin, resulting in open chromatin fibers that display greater accessibility to transcription factors and their coregulators [13]. Several studies implicate H2BK120ub in developmental processes including that optical embryonic stem cell differentiation requires dynamic changes in H2B ubiquitination patterns in a timely and well-coordinated manner [7,14].

In this study, we identified that the histone methyltransferase EZH2 methylates H2BK120 and competitively inhibits ubiquitination of this lysine residue in cancer cells. This is the first report describing identification of histone H2B as a novel substrate of EZH2 and demonstrating the biologic significance of the histone H2B methylation in human carcinogenesis.

Materials and Methods

Cell Culture

CCD-18Co, HFL1, HCT116, SW480, RT4, MCF7, HeLa, and 293T cells were from American Type Culture Collection (ATCC, Manassas, VA) in 2001 and 2003 and tested and authenticated by DNA profiling for polymorphic short tandem repeat (STR) markers. SBC5, RERF-LC-AI, and Alexander cells were from Japanese Collection of Research Bioresources (JCRB) in 2001 and tested and authenticated by DNA profiling for polymorphic STR markers.

All cell lines were grown in monolayers in the following appropriate media: Dulbecco's modified Eagle's medium for RERF-LC-AI and 293T cells; Eagle's minimum essential medium for CCD-18Co, SBC5, MCF7, and HeLa cells; Leibovitz's L-15 for SW480 cells; McCoy's 5A medium for RT4 and HCT116 cells; RPMI 1640 medium for Alexander cells; and Ham's F-12K Medium for HFL1 cells supplemented with 10% FBS and 1% antibiotic/antimycotic solution (Life Technologies, Carlsbad, CA). All cells were maintained at 37°C in humid air with 5% CO_2 condition (CCD-18Co, HFL1, SBC5, RERF-LC-AI, HCT116, SW480, Alexander, RT4, MCF7, HeLa, and 293T) or without CO_2 (SW480). Cells were transfected with FuGENE 6 (Roche Applied Science, Penzberg, Germany) according to the manufacturer's protocols [15,16].

Antibodies

The following primary antibodies were deployed: anti-FLAG [M2; Sigma-Aldrich (St Louis, MO); dilution used in Western blot (WB): 1:5000, immunocytochemistry (ICC): 1:100], anti-EZH2 (NCL-L-EZH2; Leica, Solms, Germany; dilution used in WB: 1:500), anti–histone H2B (ab1790; Abcam, Cambridge, United Kingdom; dilution used in WB: 1:1000), anti–Ubiquityl-Histone H2BK120 (No. 5546; Cell Signaling Technology, Danvers, MA; dilution used in WB: 1:1000, ICC: 1:50), and anti–actin, beta (ACTB; No. 4967; Cell Signaling Technology; dilution used in WB: 1:1000). The anti–K120-methylated H2B antibody [Sigma-Aldrich; dilution used in WB: 1:500, ICC: 1:100, immunohistochemistry (IHC): 1:5000] was produced in rabbit immunized with a synthetic peptide.

In Vitro *Methyltransferase Assay*

For the *in vitro* methyltransferase assay, recombinant Histone H2B (No. 14-410; Millipore, Billerica, MA) was incubated with EZH2 enzyme complex (No. 51004; BPS Bioscience, San Diego, CA) using 2 μCi of S-adenosyl-L-[methyl-^{3}H] methionine (Perkin-Elmer, Waltham, MA) as the methyl donor in a mixture of 10 μl of methylase activity buffer (50 mM Tris-HCl at pH8.8, 10 mM DTT, and 10 mM $MgCl_2$), for 1 hour at 30°C. Proteins were resolved on a 5% to 20% sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS-PAGE) gel (Ready Gel; Bio-Rad Laboratories, Hercules, CA) and visualized by MemCode Reversible Stain (Thermo Fisher Scientific, Waltham, MA) and fluorography.

Mass Spectrometry

Histone H2B samples reacted with EZH2 were separated on SDS-PAGE and stained with Coomassie brilliant blue. The excised Histone H2B bands were reduced in 10 mM tris(2-carboxyethyl) phosphine (Sigma-Aldrich) with 50 mM ammonium bicarbonate (Sigma-Aldrich) for 30 minutes at 37°C and alkylated in 50 mM iodoacetamide (Sigma-Aldrich) with 50 mM ammonium bicarbonate for 45 minutes in the dark at 25°C. Trypsin GOLD (Promega, Fitchburg, WI) solution was added with the enzyme-to-protein ratio at 1/50 (wt/wt) and incubated at 37°C for 16 hours. The resulting peptides were extracted from gel fragments and separated on a 0.1 × 200 mm homemade C_{18} column using 45-minute linear gradient from 2% to 35% acetonitrile in 0.1% formic acid with flow rate at 200 nl/min. The eluting peptides were analyzed with QSTAR Elite QqTOF system (AB Sciex, Framingham, MA) in the smart information-dependent acquisition mode of the Analyst QS software 2.0 (AB Sciex). The acquired mass spectrometry (MS) and tandem mass spectrometry (MS/MS) peak lists were analyzed with in-house Mascot server version

2.4.01 (Matrix Science, Boston, MA) to identify peptide sequences. We finally accepted the assigned peptides with expectation value less than 0.05 as the positive identification in Mascot Database search.

Western Blot Analysis

Samples were prepared from the cells lysed with CelLytic M Cell Lysis Reagent (Sigma-Aldrich) [17], and whole-cell lysates or immunoprecipitation (IP) products were transferred to nitrocellulose membrane. Protein bands were detected by incubating with HRP-conjugated antibodies (GE Healthcare, Little Chalfont, United Kingdom) and visualizing with Enhanced Chemiluminescence (GE Healthcare). We declare that our blots were evenly exposed in each membrane and that the blots are not clopped to the bands.

Immunocytochemistry

Cultured cells were fixed in 4% paraformaldehyde in 0.1 M phosphate buffer (pH 7.4) at room temperature for 30 minutes, permeabilized in 0.1% Triton X-100 (Sigma-Aldrich) for 3 minutes on ice, and blocked with BLOCK ACE (Yukijirushi, Sapporo, Japan) for 1 hour at room temperature. Fixed cells were incubated with primary antibodies overnight at 4°C. Then incubated with Alexa Fluor–conjugated second antibody (Life Technologies) and observed using a Leica confocal microscopy [18,19].

Immunohistochemical Analysis

The K120 monomethylation status of histone H2B in clinical tissues were examined by immunohistochemical analysis [1,17]. EnVision+ kit/HRP kit (Dako, Carpinteria, CA) was applied, and slides of paraffin-embedded lung tumor specimens were processed under high pressure (125°C, 30 seconds) in antigen-retrieval solution, high pH 9 (S2367; Dako), treated with peroxidase-blocking regent, and then treated with protein-blocking regent (K130, X0909; Dako). Tissue sections were incubated with a rabbit anti-H2BK120me1 polyclonal antibody, followed by HRP-conjugated secondary antibody (Dako). Antigen was visualized with substrate chromogen (Dako Liquid DAB Chromogen; Dako). Finally, tissue specimens were stained with Mayer haematoxylin (Hematoxylin QS; Vector Laboratories, Burlingame, CA) for 20 seconds to discriminate the nucleus from the cytoplasm. The intensity of H2BK120 methylation was evaluated using the following criteria: strong positive (scored as 2+), brown staining in >50% of tumor cells completely obscuring cytoplasm; weak positive (1+), any lesser degree of brown staining appreciable in tumor cell cytoplasm; and absent (scored as 0), no appreciable staining in tumor cells. Cases were accepted as strongly positive if two or more investigators independently defined them as such. Detailed clinical information of lung and colony tissues is described in Tables W1 and W2.

siRNA Transfection

Small interfering RNA (siRNA) oligonucleotide duplexes were purchased from Sigma-Aldrich for targeting the human *EZH2* transcript. siNegative control (siNC) was used as a control siRNA. The siRNA sequences are described in Table W3. siRNA duplexes (100 nM final concentration) were transfected into bladder and lung cancer cell lines with Lipofectamine 2000 (Life Technologies) [20].

Immunoprecipitation

Transfected 293T or SBC5 cells were lysed with CelLytic M Cell Lysis Reagent (Sigma-Aldrich) containing a complete protease inhibitor cocktail (Roche Applied Science). In a typical immunoprecipitation reaction, 300 μg of whole-cell extract was incubated with an optimum concentration of primary antibody. After the beads had been washed three times in 1 ml of TBS buffer (pH 7.6), proteins that bound to the beads were eluted by boiling in Lane Marker Reducing Sample Buffer (Thermo Fisher Scientific).

Chromatin Immunoprecipitation–Seq

Chromatin immunoprecipitation (ChIP) assays were performed using ChIP Assay kit (Millipore) according to the manufacturer's protocol. Briefly, the fragment of EZH2 and chromatin complexes was immunoprecipitated with anti-H2BK120me1 (Sigma-Aldrich) and anti-FLAG (M2; Sigma-Aldrich) antibodies 36 or 48 hours after transfection with siRNAs and pCAGGS-n3FC (Mock) or pCAGGS-n3FC-EZH2 (3× FLAG-EZH2) vectors, respectively. ChIP-DNA samples were end repaired, A tailed, and finally ligated with DNA adaptors. Subsequently, samples were size fractionated (175-225 bp) on a 2% tris-borate-EDTA (TBE) agarose gel. Fractionated ChIP-DNA samples were amplified using KAPA Library Preparation Kit (Kapa Biosystems, Woburn, MA). After a final polymerase chain reaction amplification step (18 cycles), the resulting DNA libraries were analyzed by an Agilent Technologies 2100 Bioanalyzer (Santa Clara, CA), and DNA libraries were sequenced on Illumina GAIIx (San Diego, CA).

ChIP-Seq Data Analysis

Sequencing data were acquired from the Illumina GAIIx sequencer [21]. ChIP-Seq reads were aligned using Burrows-Wheeler Aligner software (version 0.62-r126) to the human genome (University of California, Santa Cruz, CA; hg19) with default settings [22]. Burrows-Wheeler Aligner's default mismatch criteria for creating the mismatches was set up that mismatch allowance was 3%, which means that allowance of InDels (insert and deletion) was at least more than 3 bp in 100-bp reads. When same reads mapped to the same genomic position, a maximum of two reads mapping to the same position were used. For enriched-region (peak) identification (peak calling), we used the Model-based Analysis of ChIP-Seq (MACS; Harvard University, Boston, MA; version 1.4.1) algorithm [23,24]. We set up a band with 300 bp, *P* value of 1.00e-06, model fold between 10 and 30 (default) to call peaks representing enriched H2BK120me1 marks and EZH2-located regions. To prospect the peak detection of enhancer region, tags detected were aligned in each gene set within 10 kbp upstream from transcription start sites or 10 kbp downstream from transcription terminal sites. We used the statistical software DESeq (European Molecular Biology Laboratory, Hamburg, Germany), an R package, to normalize the number of tags detected in each gene region and calculated the ratio of samples and controls [25]. Normalized intensity value of genes was ordered as a heatmap using hierarchical clustering in Cluster 3.0 software (Stanford University, Stanford, CA).

Microarray Hybridization and Statistical Analysis

Microarray analysis to identify downstream genes was described previously [20,26,27]. Purified total RNA was labeled and hybridized onto Affymetrix GeneChip U133 Plus 2.0 oligonucleotide arrays (Affymetrix, Santa Clara, CA) according to the manufacturer's instructions. Probe signal intensities were normalized by robust multichip average (RMA) and Quantile (using R and Bioconductor).

ChIP Assay

ChIP assays were performed using ChIP Assay kit (17-295; Millipore) according to the manufacturer's protocol. Briefly, the fragment of K120-ubiquitinated H2B and chromatin complexes in HEK293 cells was immunoprecipitated with an anti–Ubiquityl-Histone H2BK120 antibody. After DNA fragments bound to K120-ubiquitinated H2B were eluted out, an aliquot was subjected to quantitative real-time polymerase chain reaction reactions. Protein A Agarose/Salmon Sperm DNA (16-157; Millipore) was used as a negative control. Primers were designed near the transcriptional start site of each gene, and their sequence information is shown in Table W4.

Results

EZH2 Methylates Lysine 120 on Histone H2B

To identify a novel substrate of EZH2, we conducted *in vitro* methyltransferase assays using recombinant histone proteins as well as nonhistone proteins that play important roles in human carcinogenesis. Consequently, we identified that EZH2 appeared to methylate histone H2B in addition to histone H3 (Figure W1). To validate this result, we performed *in vitro* methyltransferase assays using different amount of EZH2 enzymes and confirmed histone H2B methylation by EZH2 in a dose-dependent manner (Figure 1*A*). Subsequently, we conducted liquid chromatography (LC)-MS/MS

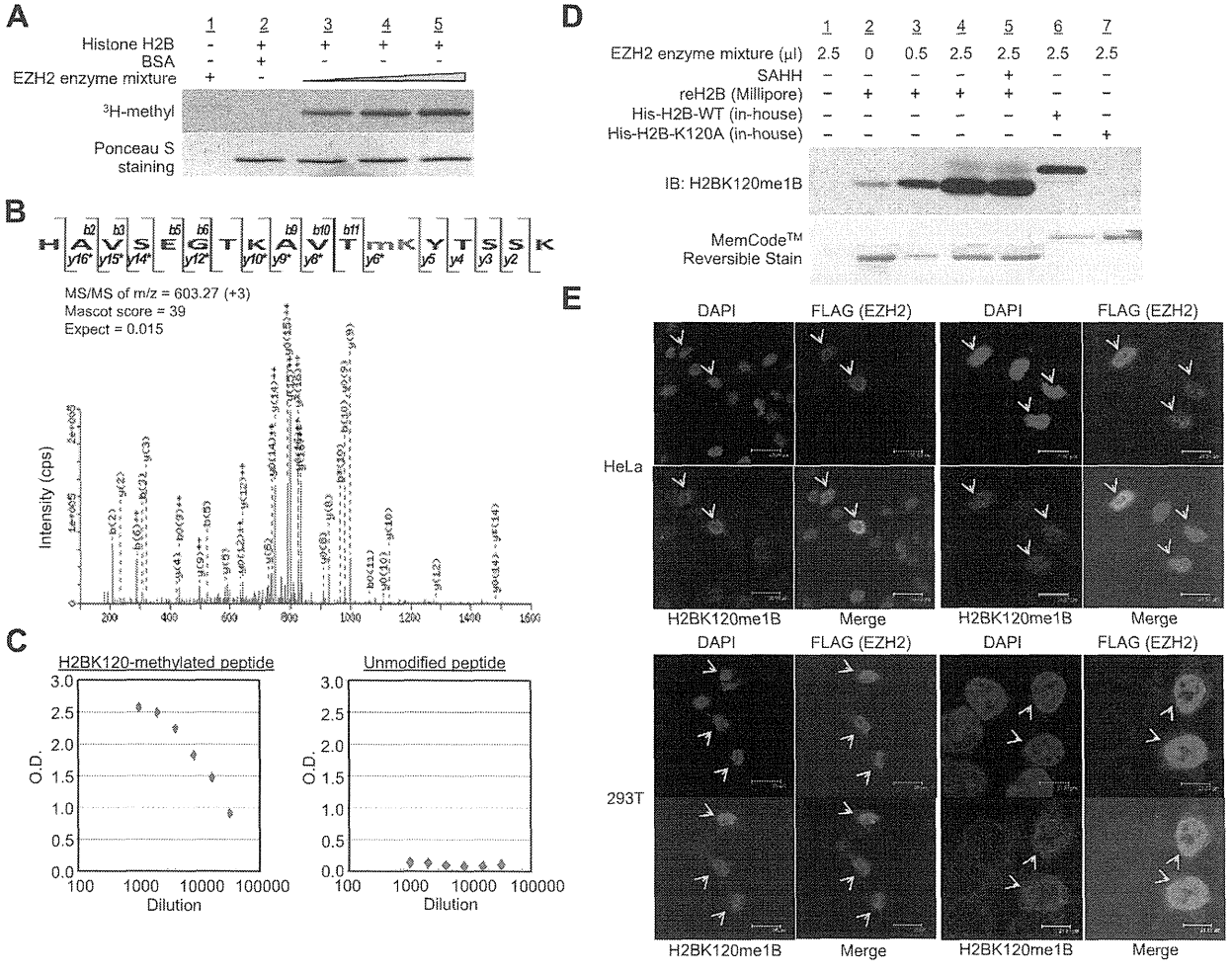

Figure 1. EZH2 methylates lysine 120 of histone H2B both *in vitro* and *in vivo*. (A) *In vitro* methyltransferase assay of H2B. Recombinant H2B and ^{3}H-SAM were incubated in the presence or absence of EZH2 enzyme complex (No. 51004; BPS Bioscience), and the reaction products were analyzed by SDS-PAGE, followed by fluorography (upper panel). The membrane was stained with Ponceau S (Sigma-Aldrich; lower panel). BSA was used as a control. (B) The MS/MS spectrum corresponding to the monomethylated H2B 109-125 peptide. The 14-Da increase of the lysine 120 residue was observed, demonstrating the monomethylated lysine 120. Score and Expect show Mascot Ion Score and Expectation value in Mascot Database search results, respectively. (C) Determination of the titer and specificity of the anti-monomethylated K120 H2B antibody analyzed by ELISA. (D) Validation of an anti-monomethylated K120 H2B antibody. Recombinant H2B-WT or H2B-K120A proteins and ^{3}H-SAM were incubated in the presence or absence of EZH2 enzyme complex, and the reaction products were analyzed by SDS-PAGE, followed by Western blot analysis with an anti-monomethylated K120 histone H2B antibody (upper panel) and stained with MemCode Reversible Stain (lower panel). SAHH, S-adenosyl-L-homocysteine hydrolase. (E) Immunocytochemical analysis of HeLa and 293T cells. Cells were stained with an anti-FLAG antibody [M2; Sigma-Aldrich; Alexa Fluor 594 (red)], an anti-H2BK120me1 antibody [Sigma-Aldrich; Alexa Fluor 488 (green)] and 4′,6′-diamidine-2′-phenylindole dihydrochloride [DAPI (blue)].