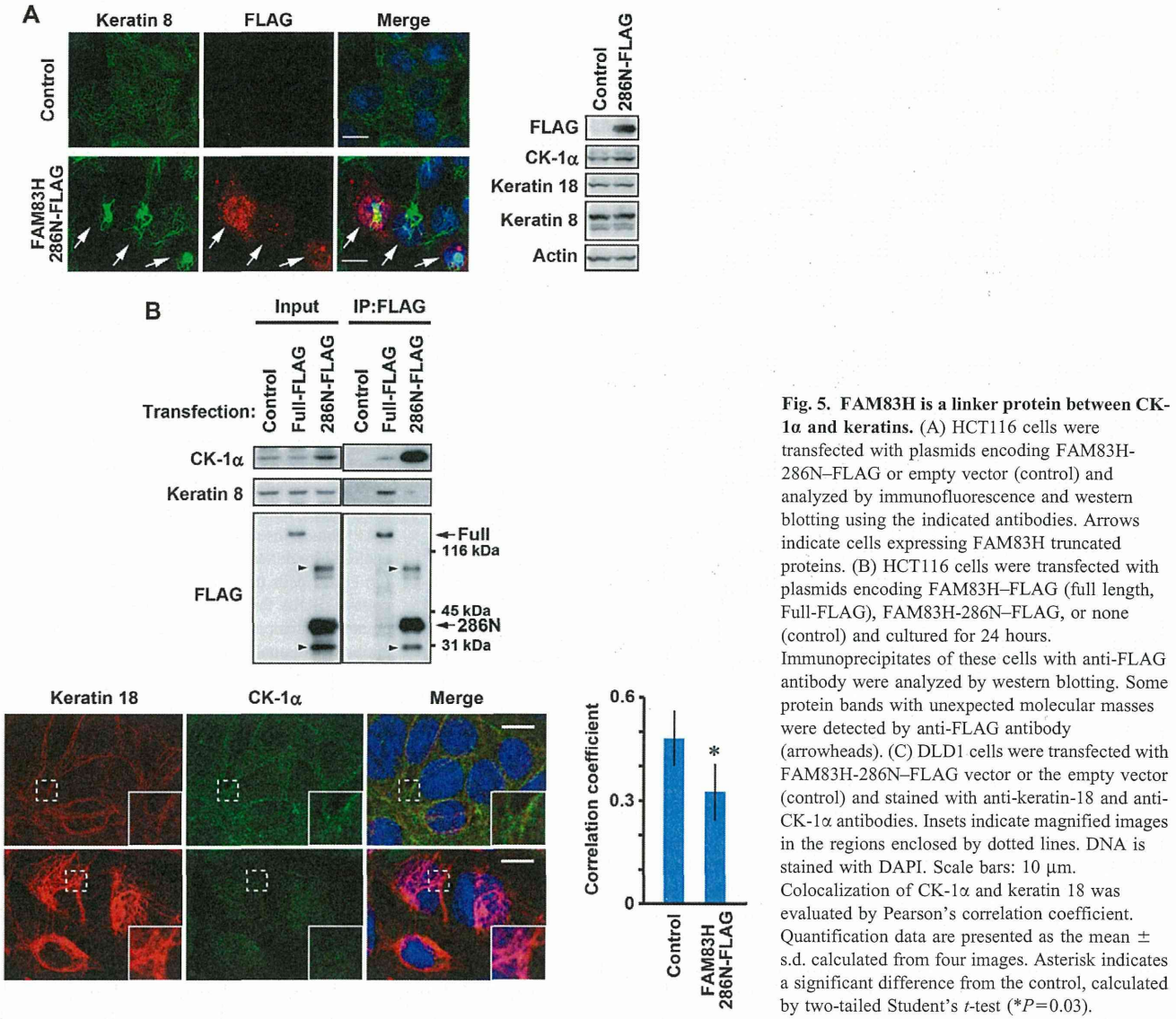


Fig. 5. FAM83H is a linker protein between CK-1α and keratins. (A) HCT116 cells were transfected with plasmids encoding FAM83H-286N–FLAG or empty vector (control) and analyzed by immunofluorescence and western blotting using the indicated antibodies. Arrows indicate cells expressing FAM83H truncated proteins. (B) HCT116 cells were transfected with plasmids encoding FAM83H–FLAG (full length, Full-FLAG), FAM83H-286N–FLAG, or none (control) and cultured for 24 hours. Immunoprecipitates of these cells with anti-FLAG antibody were analyzed by western blotting. Some protein bands with unexpected molecular masses were detected by anti-FLAG antibody (arrowheads). (C) DLD1 cells were transfected with FAM83H-286N–FLAG vector or the empty vector (control) and stained with anti-keratin-18 and anti-CK-1α antibodies. Insets indicate magnified images in the regions enclosed by dotted lines. DNA is stained with DAPI. Scale bars: 10 μm. Colocalization of CK-1α and keratin 18 was evaluated by Pearson's correlation coefficient. Quantification data are presented as the mean ± s.d. calculated from four images. Asterisk indicates a significant difference from the control, calculated by two-tailed Student's *t*-test (**P*=0.03).

epithelial cells (Fig. 7B). Note that the FAM83H expression pattern was heterogeneous throughout the tumor mass (Fig. 7C) and a subpopulation of cancer cells exhibited markedly high levels of FAM83H expression (Fig. 7C, region 1). Consistent with the *in vitro* results, FAM83H-overexpressing cancer cells exhibited disassembled keratin filaments and aberrant speckle-like colocalization of FAM83H, CK-1α and disassembled keratins (Fig. 7D,E). Another important feature of FAM83H-overexpressing cancer cells was loss or alteration of epithelial cell polarity (apical-basal polarity) (Fig. 7C, region 1) compared with FAM83H-low cancer cells (Fig. 7C, region 2). In addition, loss of E-cadherin expression, a protein that is essential for cell–cell adhesion (Birchmeier and Behrens, 1994), was also detected in FAM83H-overexpressing cancer cells (Fig. 7F). Given that the loss of epithelial cell polarity and E-cadherin expression is known to be a process during the migration or invasion of cancer cells (Thiery et al., 2009), these *in vivo* data suggest that FAM83H overexpression and the subsequent disassembly of keratin filaments occur in migrating or invading cancer cells.

FAM83H is involved in reorganization of the keratin cytoskeleton during migration of colorectal cancer cells

Reorganization of the keratin cytoskeleton is required for the migration of epithelial cells (Beil et al., 2003; Kölsch et al., 2010; Windoffer et al., 2011). To examine whether FAM83H was involved in the migration of colorectal cancer cells, we performed a wound-healing assay with FAM83H-knockdown cells. HCT116 cells treated with FAM83H siRNA or control siRNA were scratched, and the wound edge was monitored by microscopy. Treatment with FAM83H siRNA suppressed migration of the wound edge of the cell sheets to less than 50% (Fig. 8A). Visualization of the cells at the wound edge using anti-α-tubulin and anti-keratin-8 antibodies showed that control cells formed a cell-spreading structure at the wound edge and showed a redistribution of keratin filaments that was less abundant in the cell-spreading region compared with the cell body region (Fig. 8B). By contrast, FAM83H-knockdown cells formed severely bundled keratin filaments throughout the cytoplasm and were prevented from spreading outward at the wound edge

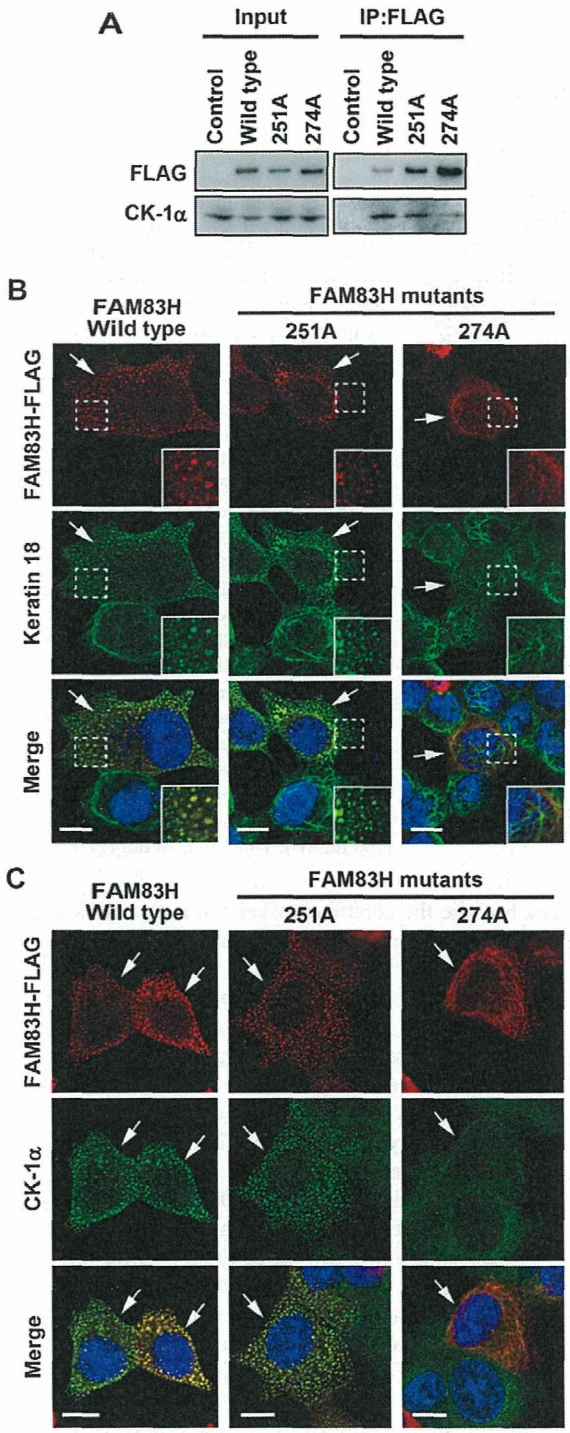


Fig. 6. A CK-1α-binding motif of FAM83H. (A) Immunoprecipitation with anti-FLAG antibody was performed using lysates from HCT116 cells transfected with wild-type FAM83H–FLAG, FAM83H-251A–FLAG or FAM83H-274A–FLAG. Input lysates and immunoprecipitates were analyzed by western blotting. (B,C) HCT116 cells were transfected with wild-type FAM83H–FLAG, FAM83H-251A–FLAG or FAM83H-274A–FLAG and analyzed by immunofluorescence with the indicated antibodies. DNA is stained with DAPI (blue). Arrows indicate cells expressing FLAG-tagged proteins. Insets indicate magnified images in the areas enclosed by dotted lines. Scale bars: 10 μm.

(Fig. 8B). These results suggest that FAM83H plays a crucial role in cell migration by regulating keratin cytoskeleton organization.

Discussion

In this study, we identified novel keratin-associated proteins, FAM83H and CK-1α, and demonstrated that FAM83H regulates keratin cytoskeleton organization by recruiting CK-1α to keratin filaments, both *in vitro* and *in vivo*. Moreover, loss of epithelial cell polarity and E-cadherin expression, in addition to keratin filament disassembly, was observed in FAM83H-overexpressing cancer cells *in vivo*. These results strongly suggest that CK-1α-mediated keratin filament disassembly induced by FAM83H overexpression is involved in the invasion and/or metastasis of colorectal cancer.

Although it has been undoubtedly accepted that the filamentous state of keratins is regulated by protein phosphorylation (Izawa and Inagaki, 2006; Omary et al., 2006; Windoffer et al., 2011), little is known about the kinases responsible for the rearrangement of keratin filaments; thus, our data showing the role of CK-1α in keratin filament rearrangement are valuable. The filamentous state of keratins is modulated by the equilibrium between assembly (bundling) and disassembly. The assembly of keratin filaments is a self-processing reaction, which does not require any additional proteins, at least *in vitro* (Steinert et al., 1976), and is highly favored over disassembly (Windoffer et al., 2011). It is therefore conceivable that the mechanism governing the disassembly of keratin filaments plays a crucial role in regulating the equilibrium of filamentous states of keratins. Aberrant speckle-like colocalization of CK-1α to keratin filaments causes disassembly of keratin filaments, whereas dissociation of CK-1α from keratin filaments causes bundling of keratin filaments. These results suggest that correct recruitment of CK-1α to keratin filaments is required to properly maintain the equilibrium of the filamentous state of keratins.

Because CK-1α-mediated disassembly of keratin filaments depends on the kinase activity of CK-1α (Fig. 3B,D), identification of the CK-1α substrates responsible for keratin filament disassembly is an important issue. As keratin proteins are known to be subject to phosphorylation to regulate their own filament status (Omary et al., 2006), keratin proteins themselves are major candidates of the CK-1α substrates; however, we could not explain the CK-1α-mediated reorganization of keratin filaments by phosphorylation, at least at Ser73 and Ser431 of keratin 8 and Ser33 and Ser52 of keratin 18 (supplementary material Fig. S2). Another candidate for CK-1α substrates is plectin, which plays a crucial role in crosslinking keratin filaments to other cellular structures (Wiche, 1998). FAM83H–FLAG immunoprecipitates contained plectin (supplementary material Table S1), and plectin knockdown, as well as the inhibition of FAM83H and CK-1α, caused keratin filament bundling in HCT116 cells (supplementary material Fig. S3) and a human liver cell line (Liu et al., 2011). Our next task will be proteomic analysis to identify the CK-1α substrates responsible for the rearrangement of keratin filaments.

CK-1α is basically constitutive active and thus its subcellular distribution is an important factor for functional regulation (Knippschild et al., 2005); however, the precise mechanism governing the subcellular distribution of CK-1α remains largely unknown. We determined that FAM83H is a novel key regulator of the subcellular distribution of CK-1α. FAM83H was found to be a linker protein between CK-1α and keratins, because

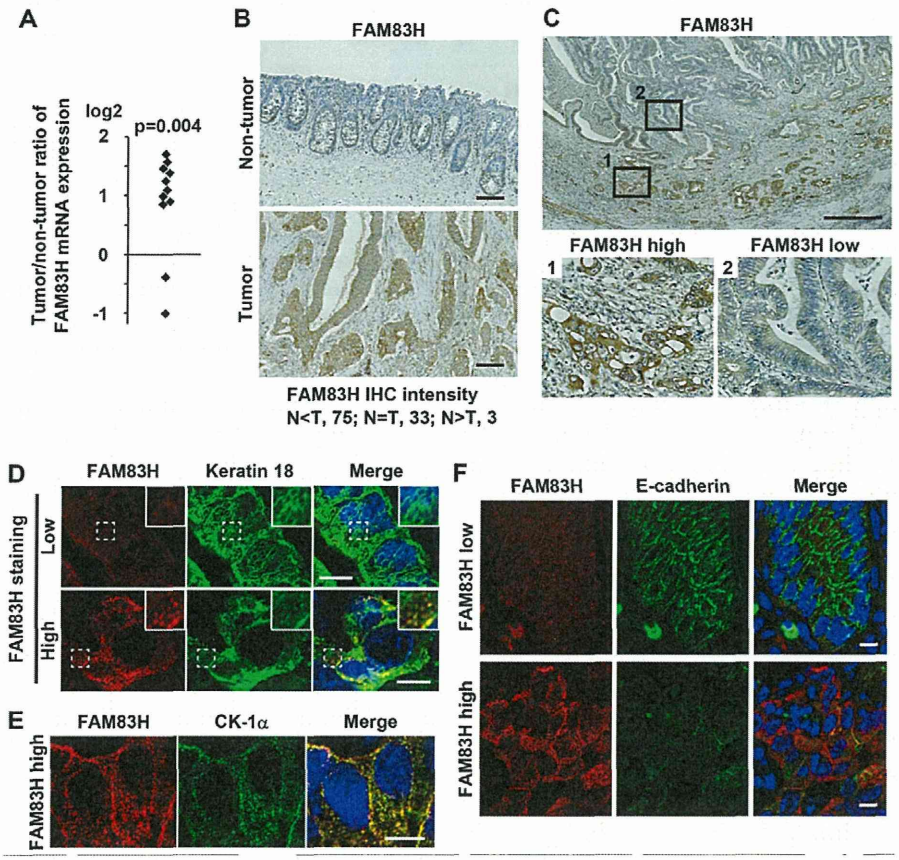


Fig. 7. Disassembly of keratin filaments and loss of epithelial cell polarity in FAM83H-overexpressing human colorectal cancer tissue. (A) Tumor and the adjacent non-tumor tissues from 12 colorectal cancer patients were analyzed by qPCR. The expression level of FAM83H in each tumor sample was normalized by that in the corresponding non-tumor sample. The values are expressed as log2 ratios. Significant differences between tumor and non-tumor samples were calculated by double-tailed Student's paired *t*-test ($P=0.004$). (B,C) Paraffin-embedded cancer tissue specimens from 111 different patients were stained with anti-FAM83H antibody. Staining intensity for FAM83H in cancer cells was high in 75 specimens, similar in 33 specimens or low in 3 specimens, compared with that in adjacent non-tumor epithelial cells. The boxed regions (regions 1 and 2) are magnified below (C). (D–F) Immunofluorescence of snap-frozen cancer tissues using the indicated antibodies. Images of cancer cells expressing high or low levels of FAM83H are shown. Insets indicate magnified images in the regions enclosed by dotted lines. DNA was stained with DAPI (blue). Scale bars: 100 μm (B); 1 mm (C); 10 μm (D–F).

FAM83H interacts with CK-1α in the N-terminal region and keratins in the C-terminal region. Consistent with the cases of NFAT1, PER1 and PER2 (Okamura et al., 2004), FAM83H interacts with CK-1α through an FxxxF amino acid sequence in the N-terminal region (FDEEFRILF, 270–278 aa).

Intriguingly, FAM83H overexpression not only induced aberrant recruitment of CK-1α to keratin filaments but also decreased cytoplasmic localization of CK-1α (Fig. 4). Cytoplasmic CK-1α plays an essential role in the phosphorylation and degradation of β-catenin (Knippschild et al., 2005). Loss of CK-1α-mediated phosphorylation of β-catenin causes nuclear accumulation of β-catenin and transcription of Wnt-specific genes responsible for the control of cell fate decisions, resulting in tumorigenesis (Elyada et al., 2011; Polakis, 2000; Valenta et al., 2012). In fact, we observed dephosphorylation and nuclear accumulation of β-catenin upon FAM83H overexpression *in vitro* and *in vivo* (unpublished data). These results suggest that the subcellular distribution and function of CK-1α are totally orchestrated by FAM83H expression levels.

Our *in vivo* data suggested the physiological significance of the mechanism governing the rearrangement of keratin filaments by FAM83H and CK-1α in colorectal cancer. FAM83H overexpression, aberrant localization of CK-1α, and keratin filament disassembly are all detected in colorectal cancer cells exhibiting loss or alteration of epithelial cell polarity. As discussed above, CK-1α is involved in Wnt–β-catenin signaling, which regulates the epithelial cell polarity of cancer cells (Thiery et al., 2009); thus, the aberrant localization of CK-1α might contribute to the loss of epithelial cell polarity. In addition, because the keratin cytoskeleton is an essential element for maintenance of epithelial cell polarity (Ameen et al., 2001; Oriolo et al., 2007; Salas et al., 1997), keratin filament disassembly caused by FAM83H overexpression may also occur during the loss of epithelial cell polarity.

E-cadherin expression is also suppressed in FAM83H-overexpressing cancer cells *in vivo*. Loss of E-cadherin expression, as well as that of epithelial cell polarity, is a hallmark of epithelial–mesenchymal transition (EMT), which is a process by which cancer cells escape from the primary tumor mass for invasion and metastasis (Thiery et al., 2009). Moreover, E-cadherin expression is controlled by Wnt/β-catenin signaling (Thiery et al., 2009). These results imply that the aberrant localization of CK-1α caused by FAM83H overexpression might be involved in the suppression of E-cadherin and EMT of colorectal cancer cells.

Several reports have suggested that expression levels of keratin proteins are correlated to the cancer grade and patient survival (Knösel et al., 2006; Moll et al., 2008) and that keratin 18 contributes to suppression of the invasiveness of breast and pancreatic cancer cells (Bühler and Schaller, 2005; Pankov et al., 1997). In addition, Beil and colleagues proposed that reorganization of the keratin cytoskeleton by sphingosylphosphorylcholine induced cellular elasticity and enhanced cell migration of pancreatic cancer cells (Beil et al., 2003). Given our *in vitro* data suggesting the role of FAM83H-mediated rearrangement of keratin filaments in the migration of colorectal cancer cells, these results suggest that the disassembly of keratin filaments by overexpression of FAM83H contributes to the invasion and metastasis of colorectal cancer.

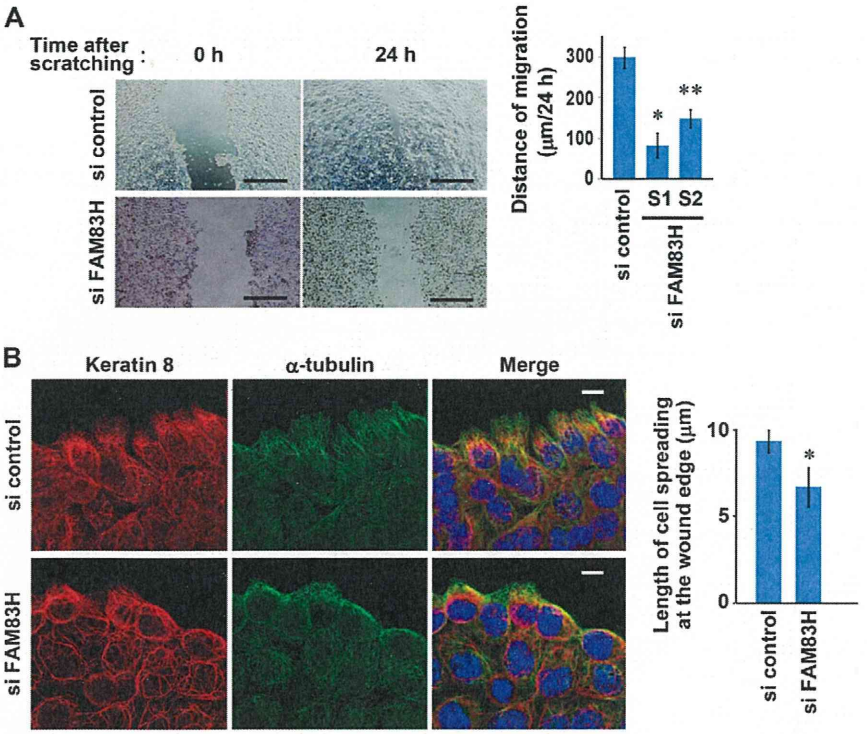


Fig. 8. FAM83H knockdown inhibits reorganization of the keratin cytoskeleton during migration of colorectal cancer cells. (A) HCT116 cells were treated with FAM83H siRNA or control siRNA for 24 hours, and then the cell sheets were scratched and cultured for 24 hours. Representative images of the cell sheets at 0 and 24 hours after scratching are shown. The migration distance of the wound edge of the cell sheets during 24 hours of culture was measured. Quantification data are presented as the mean ± s.d. of three experiments. (B) HCT116 cells were treated with FAM83H siRNA or control siRNA for 36 hours and the cell sheets were scratched. The cells were then cultured for 6 hours and stained with anti-keratin-8 and anti-α-tubulin antibodies, and DAPI (blue). Confocal images of representative cells at the wound edge are shown. The extent of spread of each leading cell at the wound edge was assessed by measuring the distance from the nucleus to the cell periphery. Quantification data are presented as the mean ± s.d. of three experiments. Approximately 100 leading cells were assessed in each experiment. Asterisks indicate significant differences from the control, calculated by two-tailed Student's *t*-test (**P*=0.0007, ***P*=0.001 in A; **P*=0.02 in B). Scale bars: 500 μm (A); 10 μm (B).

In conclusion, this work elucidated a novel mechanism governing the rearrangement of keratin filaments in colorectal cancer cells. Our data also suggested that aberrant localization of CK-1α and keratin filament disassembly induced by FAM83H overexpression are involved in the migration of colorectal cancer cells. FAM83H overexpression was suggested in various types of cancer (Sasaroli et al., 2011), and CK-1α is known to play a tumor suppressive role (Elyada et al., 2011; Valenta et al., 2012). Thus, tumor progression of various types of cancer might be commonly promoted by FAM83H overexpression, malfunction of CK-1α and subsequent disassembly of keratin filaments.

Materials and Methods

Cell culture and transfection

HCT116 and DLD1 colorectal cancer cell lines were purchased from the American Type Culture Collection (ATCC, Manassas, VA, USA). Cells were grown at 37℃ in 5% CO_2 in Iscove's Modified Dulbecco's Medium (IMDM; Invitrogen, Carlsbad, CA) supplemented with 10% fetal bovine serum (FBS; Invitrogen). Plasmids and siRNA were transfected using Lipofectamine 2000 (Invitrogen) and Lipofectamine RNAiMAX (Invitrogen), respectively. D4476, a CK-1 inhibitor, was used at a concentration of 100 μM (Abcam, Cambridge, UK).

Plasmid DNA and siRNA

To generate FAM83H–FLAG-expressing plasmid, cDNA encoding human *FAM83H* (NM_198488.3) was amplified from HCT116 cells by PCR using the forward primer 5′-ATAGAATTCAACATGGCCCGTCGCTCTCAGAG-3′ and the reverse primer 5′-ACGGGATCCTCCCTTCTTGCTTTTGAACG-3′ and cloned into the p3XFLAG-CMV-14 vector (Sigma-Aldrich, St Louis, MO). *FAM83H* cDNA from HCT116 cells has three silent mutations (1896C/T, 2001C/A and 2953C/T). To generate plasmids expressing FLAG-tagged N-terminal fragments of FAM83H (amino acids 1–286, FAM83H-286N–FLAG; 1–296, FAM83H-296N–FLAG), the corresponding regions of *FAM83H* cDNA were amplified by PCR using the common forward primer 5′-ATAGAATTCA-ACATGGCCCGTCGCTCTCAGAG-3′ and the reverse primer 5′-ATAG-GATCCGGGCACAAGCGGCTCGGACTG-3′ for FAM83H-286N–FLAG and 5′-ATAGGATCCGGCGTCCATGCGGGCCAG-3′ for FAM83H-296N–FLAG and cloned into the p3XFLAG-CMV-14 vector. FAM83H-251A–FLAG and FAM83H-274A–FLAG were generated using a PrimeSTAR Mutagenesis Basal Kit (Takara Bio, Shiga, Japan) and FAM83H–FLAG vector. For the 251A mutant, 5′-TGGTCCGCTGAGAAGATCCACCGCAGC-3′ and 5′-CTTCTCAGCGGA-CCACATGAAGCTGTA-3′ were used as PCR primers. For the 274A mutant, 5′-GAGGAGGCCCGCATCCTCTTCGCGCAG-3′ and 5′-GATGCGGGCCT-CCTCGTCGAAGCTGGA-3′ were used as PCR primers. FAM83H siRNAs (FAM83H siRNA S1, FAM83H-HSS138852; FAM83H siRNA S2, FAM83H-HSS138851) were purchased from Invitrogen. CK-1α siRNAs targeting the sequence 5′-CAGAATTTGCGATGTACTT-3′ or 5′-GAATTTGCGATGTACTT-AA-3′ and plectin siRNA targeting the sequence 5′-CCAAGAACTTGCA-GAAGTT-3′ were purchased from Sigma-Aldrich. The control siRNA duplexes were purchased from Invitrogen (Medium GC Duplex #2) and Sigma-Aldrich (Mission negative control SIC-001).

Antibodies

The following antibodies were purchased: anti-FAM83H (HPA024604; Sigma-Aldrich), anti-keratin-8 (TS1, Thermo Scientific, Fremont, CA; EP1628Y, Epitomics, Burlingame, CA), anti-keratin-18 (DC10; Thermo Scientific), anti-keratin 19 (RCK108; Thermo Scientific), anti-α-tubulin (DM1A, Sigma-Aldrich; YOL1/34, Santa Cruz Biotechnology, Santa Cruz, CA), anti-FLAG (M2; Sigma-Aldrich), anti-E-cadherin (36B5; Thermo Scientific), anti-CK-1α (C-19; Santa Cruz Biotechnology), anti-phospho-keratin-8-Ser431 (EP1630; Epitomics), anti-phospho-keratin-8-Ser73 (E431-2; Abcam), anti-phospho-keratin-18-Ser33 (sc-101727; Santa Cruz Biotechnology), anti-phospho-keratin-18-Ser52 (sc-17032; Santa Cruz Biotechnology), anti-plectin (sc-7572; Santa Cruz Biotechnology) and anti-actin (C-11; Santa Cruz Biotechnology). Mouse IgG1 control (MOPC-21; Exbio, Vestec, Czech Republic) was used for immunoprecipitation. Alexa Fluor 488 and Alexa Fluor 594 donkey anti-mouse IgG and Alexa Fluor 488 and Alexa Fluor 594 donkey anti-rabbit IgG, Alexa Fluor 488 donkey anti-goat IgG, and Alexa Fluor 488 donkey anti-rat IgG antibodies were used for immunofluorescence (Invitrogen). HRP-conjugated horse anti-mouse IgG (Cell Signaling Technology, Beverly, MA), donkey anti-rabbit IgG (GE Healthcare, Little Chalfont, UK), and donkey anti-goat IgG (Santa Cruz Biotechnology) antibodies were used for western blotting.

Tissue samples from colorectal cancer patients

Tumor and adjacent non-tumor tissue samples were collected from patients with colorectal cancer in the Department of Frontier Surgery, Chiba University Hospital. For qPCR, tissue samples were rapidly frozen and stored at −80℃. For immunofluorescence, tissue samples were embedded in OCT compound (Sakura Finetek, Tokyo, Japan), snap frozen in liquid nitrogen, and cut into 5 μm sections using a cryostat (Hyrax C50; Carl Zeiss, Jena, Germany). Paraffin-embedded blocks were cut into 2.5 μm sections using a microtome, REM-710 (Yamato, Saitama, Japan). The protocol for the collection and use of the tissue samples was

approved by the ethics committees of the Graduate School of Medicine, Chiba University and the Proteome Research Center, National Institute of Biomedical Innovation. Written informed consent was obtained from each patient before surgery.

Quantitative PCR

Extraction of total RNA and cDNA synthesis were performed as described previously (Tomonaga et al., 2004). qPCR reaction was performed using Power CYBR Green reagents (Applied Biosystems, Foster City, CA). The levels of mRNA encoding FAM83H and β-actin were examined, and the value of FAM83H was normalized by that of β-actin in each tissue sample. The following primer pairs were used: FAM83H, forward primer 5′-cgacaagtgccgtgtcaacc-3′ and reverse primer 5′-acttcccagtgcggcagtag-3′; β-actin, forward primer 5′-agaaaatctggcaccacacc-3′ and reverse primer 5′-ggggtgttgaaggtctcaaa-3′.

Protein extraction, immunoprecipitation and western blotting

For the extraction of whole cellular proteins of cell lines, cells were directly lysed in SDS-PAGE sample buffer. For preparation of cell lysates used for immunoprecipitation (IP lysates), cells were suspended in PBS containing 1% NP40, Complete protease inhibitor cocktail (Roche, Basel, Switzerland), and PhosSTOP phosphatase inhibitor cocktail (Roche), and then homogenized by sonication. After centrifugation at 100,000 ***g*** for 30 minutes, the supernatant was collected. Immunoprecipitation was performed using antibodies crosslinked to Protein G Dynabeads (Invitrogen) using dimethyl pimelimidate dehydrochloride (MP Biochemicals, Santa Ana, CA). IP lysates were reacted with antibody-coated Dynabeads for 1 hour at 4℃, and the absorbed proteins were eluted with 100 mM glycine-HCl (pH 3.0) or SDS-PAGE sample buffer. Western blotting was performed using the chemiluminescence detection system ECL or ECL Prime (GE Healthcare). The images were obtained with LAS4000 (Fuji Film, Tokyo, Japan) and processed with Multi Gauge V3.2 (Fuji Film) and Photoshop CS5 (Adobe, San Jose, CA).

Protein identification by LC-MS/MS

Immunoprecipitates with anti-FLAG antibody were resolved by SDS-PAGE and the gel lane was divided into nine pieces corresponding to different molecular masses. In-gel tryptic digestion of proteins was performed as described previously (Adachi et al., 2007). The digested peptides were analyzed using an LTQ-Orbitrap Velos mass spectrometer (Thermo Scientific) equipped with a nanoHPLC system (Paradigm MS2; Michrom, Auburn, CA) and an HTC-PAL autosampler (Zwingen, Kanton Bern, Switzerland). A 0.3×5 mm trap column [L-column ODS; Chemicals Evaluation and Research Institute (CERI), Tokyo, Japan] and an analytical column made in-house by packing L-column2 C18 (CERI) into a self-pulled needle (0.1×200 mm) were used (Muraoka et al., 2012; Narumi et al., 2012). The mobile phases consisted of buffer A (0.1% formic acid and 2% acetonitrile) and B (0.1% formic acid and 90% acetonitrile). The nanoLC gradient was delivered at 500 nl/minute and consisted of a linear gradient of buffer B developed from 5 to 35% B in 45 minutes. The dynamic exclusion function of LTQ-Orbitrap was turned off. For protein identification, peptide mass data were matched by searching the UniProtKB/Swiss-Prot database (2011_12) using the MASCOT search engine v2.3. Database search parameters were: the charge of the precursor ion, 2+ and 3+; peptide mass tolerance, 3 ppm; fragment tolerance, 0.6 Da; allowing up to one missed cleavage; fixed modification, carbamidomethylation of cysteine; variable modification, oxidation of methionine. Proteins were identified based on at least two unique peptides. The number of assigned spectra was calculated using Scaffold 3 software (Proteome Software, Portland, OR) for semi-quantitation.

Immunofluorescence and immunohistochemistry

Cells or tissue sections from snap-frozen tissues were fixed with methanol at −20℃ for 2 minutes or 4% paraformaldehyde in PBS at 30℃ for 20 minutes, permeabilized with 0.5% Triton X-100 in PBS for 5 minutes on ice, blocked in PBS containing 0.01% Tween 20 and 3% BSA on ice, and sequentially incubated with primary and secondary antibodies at room temperature. For immunofluorescence, DNA was stained with 100 ng/ml of 4′-6-diamidino-2-phenylindole (DAPI; Sigma-Aldrich) and stained samples were viewed under an LSM710 confocal microscope with Zen software (Carl Zeiss, Jena, Germany). The objective lenses were EC Plan-NEO FLUAR 10× /0.3 or 40× /1.3 and Plan APOCHROMAT 63× /1.4. Pearson's correlation coefficient was calculated using Zen software, according to the manufacturer's instructions. The value can range from 1 to −1, with 1 indicating a complete positive correlation, −1 a negative correlation and with 0 indicating no correlation. For immunohistochemistry, visualization was performed with diaminobenzidine (DAB) chromogen (EnVision+ Kit/HRP; Dako, Glostrup, Denmark) and nuclei were stained with Mayer's hematoxylin (Muto Pure Chemicals, Tokyo, Japan) (Seimiya et al., 2008). Stained sections were scanned on a NanoZoomer RS digital slide imaging system (Hamamatsu Photonics, Hamamatsu, Japan). Two pathologists evaluated the immunostaining of colorectal cancer tissues. Composite figures were prepared using Photoshop CS5.

Acknowledgements

We would like to thank Hiromi Saito, Masayoshi Kuwano, Shio Watanabe and Yuuki Hashimoto for technical assistance.

Author contributions

T.K. designed the study, performed experiments and prepared manuscript; T.T. designed the study and prepared the manuscript; H.K. contributed to the design of the study; I.H., T.N. and H.M. collected human tissue samples; N.K. gave technical support; J.A., M.S. and T.S. gave advice on the design of proteomic, pathological and biological experiments.

Funding

This work was supported by a Grant-in-Aid, Research on Biological Markers for New Drug Development [grant number H20-0005 to T.T.]; the Ministry of Health, Labour and Grants-in-Aids [grant numbers 21390354 to T.T. and 22790539 to T.K.]; the Ministry of Education, Science, Sports and Culture of Japan.

Supplementary material available online at
http://jcs.biologists.org/lookup/suppl/doi:10.1242/jcs.129684/-/DC1

References

Adachi, J., Kumar, C., Zhang, Y. and Mann, M. (2007). In-depth analysis of the adipocyte proteome by mass spectrometry and bioinformatics. *Mol. Cell. Proteomics* **6**, 1257-1273.

Ameen, N. A., Figueroa, Y. and Salas, P. J. (2001). Anomalous apical plasma membrane phenotype in CK8-deficient mice indicates a novel role for intermediate filaments in the polarization of simple epithelia. *J. Cell Sci.* **114**, 563-575.

Beil, M., Micoulet, A., von Wichert, G., Paschke, S., Walther, P., Omary, M. B., Van Veldhoven, P. P., Gern, U., Wolff-Hieber, E., Eggermann, J. et al. (2003). Sphingosylphosphorylcholine regulates keratin network architecture and visco-elastic properties of human cancer cells. *Nat. Cell Biol.* **5**, 803-811.

Birchmeier, W. and Behrens, J. (1994). Cadherin expression in carcinomas: role in the formation of cell junctions and the prevention of invasiveness. *Biochim. Biophys. Acta* **1198**, 11-26.

Boczonadi, V., McInroy, L. and Määttä, A. (2007). Cytolinker cross-talk: periplakin N-terminus interacts with plectin to regulate keratin organisation and epithelial migration. *Exp. Cell Res.* **313**, 3579-3591.

Bühler, H. and Schaller, G. (2005). Transfection of keratin 18 gene in human breast cancer cells causes induction of adhesion proteins and dramatic regression of malignancy in vitro and in vivo. *Mol. Cancer Res.* **3**, 365-371.

Elyada, E., Pribluda, A., Goldstein, R. E., Morgenstern, Y., Brachya, G., Cojocaru, G., Snir-Alkalay, I., Burstain, I., Haffner-Krausz, R., Jung, S. et al. (2011). CKIα ablation highlights a critical role for p53 in invasiveness control. *Nature* **470**, 409-413.

Flitney, E. W., Kuczmarski, E. R., Adam, S. A. and Goldman, R. D. (2009). Insights into the mechanical properties of epithelial cells: the effects of shear stress on the assembly and remodeling of keratin intermediate filaments. *FASEB J.* **23**, 2110-2119.

Hatanaka, H., Takada, S., Tsukui, M., Choi, Y. L., Kurashina, K., Soda, M., Yamashita, Y., Haruta, H., Hamada, T., Tamada, K. et al. (2010). Identification of the transforming activity of Indian hedgehog by retroviral expression screening. *Cancer Sci.* **101**, 60-64.

Izawa, I. and Inagaki, M. (2006). Regulatory mechanisms and functions of intermediate filaments: a study using site- and phosphorylation state-specific antibodies. *Cancer Sci.* **97**, 167-174.

Kim, J. W., Lee, S. K., Lee, Z. H., Park, J. C., Lee, K. E., Lee, M. H., Park, J. T., Seo, B. M., Hu, J. C. and Simmer, J. P. (2008). FAM83H mutations in families with autosomal-dominant hypocalcified amelogenesis imperfecta. *Am. J. Hum. Genet.* **82**, 489-494.

Klammer, M., Kaminski, M., Zedler, A., Oppermann, F., Blencke, S., Marx, S., Müller, S., Tebbe, A., Godl, K. and Schaab, C. (2012). Phosphosignature predicts dasatinib response in non-small cell lung cancer. *Mol. Cell. Proteomics* **11**, 651-668.

Knippschild, U., Gocht, A., Wolff, S., Huber, N., Löhler, J. and Stöter, M. (2005). The casein kinase 1 family: participation in multiple cellular processes in eukaryotes. *Cell. Signal.* **17**, 675-689.

Knösel, T., Emde, V., Schlüns, K., Schlag, P. M., Dietel, M. and Petersen, I. (2006). Cytokeratin profiles identify diagnostic signatures in colorectal cancer using multiplex analysis of tissue microarrays. *Cell. Oncol.* **28**, 167-175.

Kölsch, A., Windoffer, R., Würflinger, T., Aach, T. and Leube, R. E. (2010). The keratin-filament cycle of assembly and disassembly. *J. Cell Sci.* **123**, 2266-2272.

Lee, S. K., Hu, J. C., Bartlett, J. D., Lee, K. E., Lin, B. P., Simmer, J. P. and Kim, J. W. (2008). Mutational spectrum of FAM83H: the C-terminal portion is required for tooth enamel calcification. *Hum. Mutat.* **29**, E95-E99.

Liu, Y. H., Cheng, C. C., Ho, C. C., Chao, W. T., Pei, R. J., Hsu, Y. H., Ho, L. C., Shiu, B. H. and Lai, Y. S. (2011). Plectin deficiency on cytoskeletal disorganization and transformation of human liver cells in vitro. *Med. Mol. Morphol.* **44**, 21-26.

Long, H. A., Boczonadi, V., McInroy, L., Goldberg, M. and Määttä, A. (2006). Periplakin-dependent re-organisation of keratin cytoskeleton and loss of collective migration in keratin-8-downregulated epithelial sheets. *J. Cell Sci.* **119**, 5147-5159.
Miller, R. K., Vikstrom, K. and Goldman, R. D. (1991). Keratin incorporation into intermediate filament networks is a rapid process. *J. Cell Biol.* **113**, 843-855.
Miller, R. K., Khuon, S. and Goldman, R. D. (1993). Dynamics of keratin assembly: exogenous type I keratin rapidly associates with type II keratin in vivo. *J. Cell Biol.* **122**, 123-135.
Moll, R., Divo, M. and Langbein, L. (2008). The human keratins: biology and pathology. *Histochem. Cell Biol.* **129**, 705-733.
Muraoka, S., Kume, H., Watanabe, S., Adachi, J., Kuwano, M., Sato, M., Kawasaki, N., Kodera, Y., Ishitobi, M., Inaji, H. et al. (2012). Strategy for SRM-based verification of biomarker candidates discovered by iTRAQ method in limited breast cancer tissue samples. *J. Proteome Res.* **11**, 4201-4210.
Narumi, R., Murakami, T., Kuga, T., Adachi, J., Shiromizu, T., Muraoka, S., Kume, H., Kodera, Y., Matsumoto, M., Nakayama, K. et al. (2012). A strategy for large-scale phosphoproteomics and SRM-based validation of human breast cancer tissue samples. *J. Proteome Res.* **11**, 5311-5322.
Okamura, H., Garcia-Rodriguez, C., Martinson, H., Qin, J., Virshup, D. M. and Rao, A. (2004). A conserved docking motif for CK1 binding controls the nuclear localization of NFAT1. *Mol. Cell. Biol.* **24**, 4184-4195.
Omary, M. B., Ku, N. O., Tao, G. Z., Toivola, D. M. and Liao, J. (2006). "Heads and tails" of intermediate filament phosphorylation: multiple sites and functional insights. *Trends Biochem. Sci.* **31**, 383-394.
Omary, M. B., Ku, N. O., Strnad, P. and Hanada, S. (2009). Toward unraveling the complexity of simple epithelial keratins in human disease. *J. Clin. Invest.* **119**, 1794-1805.
Oriolo, A. S., Wald, F. A., Ramsauer, V. P. and Salas, P. J. (2007). Intermediate filaments: a role in epithelial polarity. *Exp. Cell Res.* **313**, 2255-2264.
Osmanagic-Myers, S., Gregor, M., Walko, G., Burgstaller, G., Reipert, S. and Wiche, G. (2006). Plectin-controlled keratin cytoarchitecture affects MAP kinases involved in cellular stress response and migration. *J. Cell Biol.* **174**, 557-568.
Pankov, R., Simcha, I., Zöller, M., Oshima, R. G. and Ben-Ze'ev, A. (1997). Contrasting effects of K8 and K18 on stabilizing K19 expression, cell motility and tumorigenicity in the BSp73 adenocarcinoma. *J. Cell Sci.* **110**, 965-974.
Polakis, P. (2000). Wnt signaling and cancer. *Genes Dev.* **14**, 1837-1851.
Salas, P. J., Rodriguez, M. L., Viciana, A. L., Vega-Salas, D. E. and Hauri, H. P. (1997). The apical submembrane cytoskeleton participates in the organization of the apical pole in epithelial cells. *J. Cell Biol.* **137**, 359-375.
Sasaroli, D., Gimotty, P. A., Pathak, H. B., Hammond, R., Kougioumtzidou, E., Katsaros, D., Buckanovich, R., Devarajan, K., Sandaltzopoulos, R., Godwin, A. K. et al. (2011). Novel surface targets and serum biomarkers from the ovarian cancer vasculature. *Cancer Biol. Ther.* **12**, 169-180.
Seimiya, M., Tomonaga, T., Matsushita, K., Sunaga, M., Oh-Ishi, M., Kodera, Y., Maeda, T., Takano, S., Togawa, A., Yoshitomi, H. et al. (2008). Identification of novel immunohistochemical tumor markers for primary hepatocellular carcinoma; clathrin heavy chain and formiminotransferase cyclodeaminase. *Hepatology* **48**, 519-530.
Sivaramakrishnan, S., Schneider, J. L., Sitikov, A., Goldman, R. D. and Ridge, K. M. (2009). Shear stress induced reorganization of the keratin intermediate filament network requires phosphorylation by protein kinase C zeta. *Mol. Biol. Cell* **20**, 2755-2765.
Steinert, P. M., Idler, W. W. and Zimmerman, S. B. (1976). Self-assembly of bovine epidermal keratin filaments in vitro. *J. Mol. Biol.* **108**, 547-567.
Strnad, P., Windoffer, R. and Leube, R. E. (2001). In vivo detection of cytokeratin filament network breakdown in cells treated with the phosphatase inhibitor okadaic acid. *Cell Tissue Res.* **306**, 277-293.
Strnad, P., Windoffer, R. and Leube, R. E. (2002). Induction of rapid and reversible cytokeratin filament network remodeling by inhibition of tyrosine phosphatases. *J. Cell Sci.* **115**, 4133-4148.
Thiery, J. P., Acloque, H., Huang, R. Y. and Nieto, M. A. (2009). Epithelial-mesenchymal transitions in development and disease. *Cell* **139**, 871-890.
Tomonaga, T., Matsushita, K., Yamaguchi, S., Oh-Ishi, M., Kodera, Y., Maeda, T., Shimada, H., Ochiai, T. and Nomura, F. (2004). Identification of altered protein expression and post-translational modifications in primary colorectal cancer by using agarose two-dimensional gel electrophoresis. *Clin. Cancer Res.* **10**, 2007-2014.
Urzúa, B., Ortega-Pinto, A., Morales-Bozo, I., Rojas-Alcayaga, G. and Cifuentes, V. (2011). Defining a new candidate gene for amelogenesis imperfecta: from molecular genetics to biochemistry. *Biochem. Genet.* **49**, 104-121.
Valenta, T., Hausmann, G. and Basler, K. (2012). The many faces and functions of β-catenin. *EMBO J.* **31**, 2714-2736.
Wiche, G. (1998). Role of plectin in cytoskeleton organization and dynamics. *J. Cell Sci.* **111**, 2477-2486.
Windoffer, R. and Leube, R. E. (1999). Detection of cytokeratin dynamics by time-lapse fluorescence microscopy in living cells. *J. Cell Sci.* **112**, 4521-4534.
Windoffer, R., Wöll, S., Strnad, P. and Leube, R. E. (2004). Identification of novel principles of keratin filament network turnover in living cells. *Mol. Biol. Cell* **15**, 2436-2448.
Windoffer, R., Beil, M., Magin, T. M. and Leube, R. E. (2011). Cytoskeleton in motion: the dynamics of keratin intermediate filaments in epithelia. *J. Cell Biol.* **194**, 669-678.
Wöll, S., Windoffer, R. and Leube, R. E. (2007). p38 MAPK-dependent shaping of the keratin cytoskeleton in cultured cells. *J. Cell Biol.* **177**, 795-807.
Wright, J. T., Frazier-Bowers, S., Simmons, D., Alexander, K., Crawford, P., Han, S. T., Hart, P. S. and Hart, T. C. (2009). Phenotypic variation in FAM83H-associated amelogenesis imperfecta. *J. Dent. Res.* **88**, 356-360.
Yano, T., Tokui, T., Nishi, Y., Nishizawa, K., Shibata, M., Kikuchi, K., Tsuiki, S., Yamauchi, T. and Inagaki, M. (1991). Phosphorylation of keratin intermediate filaments by protein kinase C, by calmodulin-dependent protein kinase and by cAMP-dependent protein kinase. *Eur. J. Biochem.* **197**, 281-290.

Journal of Cell Science

Journal of proteome research

Article

pubs.acs.org/jpr

Identification of Missing Proteins in the neXtProt Database and Unregistered Phosphopeptides in the PhosphoSitePlus Database As Part of the Chromosome-Centric Human Proteome Project

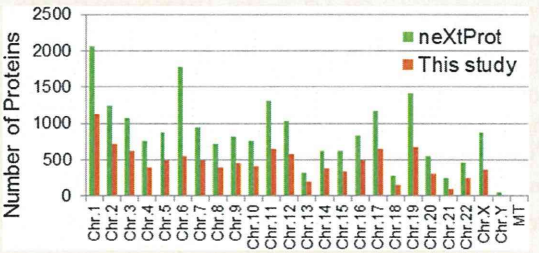
Takashi Shiromizu,[†] Jun Adachi,[†] Shio Watanabe,[†] Tatsuo Murakami,[†] Takahisa Kuga,[†] Satoshi Muraoka,[†] and Takeshi Tomonaga*[,†,‡]

[†]Laboratory of Proteome Research, National Institute of Biomedical Innovation, Ibaraki, Osaka, Japan

[‡]Clinical Proteomics Research Center, Chiba University Hospital, Chiba, Japan

Supporting Information

ABSTRACT: The Chromosome-Centric Human Proteome Project (C-HPP) is an international effort for creating an annotated proteomic catalog for each chromosome. The first step of the C-HPP project is to find evidence of expression of all proteins encoded on each chromosome. C-HPP also prioritizes particular protein subsets, such as those with post-translational modifications (PTMs) and those found in low abundance. As participants in C-HPP, we integrated proteomic and phosphoproteomic analysis results from chromosome-independent biomarker discovery research to create a chromosome-based list of proteins and phosphorylation sites. Data were integrated from five independent colorectal cancer (CRC) samples (three types of clinical tissue and two types of cell lines) and lead to the identification of 11,278 proteins, including 8,305 phosphoproteins and 28,205 phosphorylation sites; all of these were categorized on a chromosome-by-chromosome basis. In total, 3,033 "missing proteins", i.e., proteins that currently lack evidence by mass spectrometry, in the neXtProt database and 12,852 unknown phosphorylation sites not registered in the PhosphoSitePlus database were identified. Our in-depth phosphoproteomic study represents a significant contribution to C-HPP. The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium with the data set identifier PXD000089



KEYWORDS: *Chromosome-Centric Human Proteome Project, missing protein, phosphopeptide, IMAC, colorectal cancer, FASP, neXtProt, PhosphoSitePlus*

INTRODUCTION

The Chromosome-Centric Human Proteome Project (C-HPP) is a worldwide effort by proteomics researchers to create expression profiles of the approximately 20,000 genes encoded on all human chromosomes and build a database.[1] Protein expression patterns are closely associated with the location of the gene on a chromosome and are correlated with diseases associated with chromosomal abnormalities. Therefore, a comprehensive understanding of the protein expression profile of each chromosome is critical for biological studies and clinical research. The initial aim of C-HPP was to identify at least one protein isoform for every gene encoded by the human genome. Proteins not detected by antibody or proteomic analysis using mass spectrometry are called "missing proteins".[2] Currently, there are about 6,000 missing proteins among all of the proteins in the neXtProt database.[3] One reason why missing proteins are undetectable is that protein expression differs significantly between tissue and cell types. Although the number of proteins that can be identified in a single analysis has greatly increased due to recent advances in mass spectrometric techniques, complete expression profiles of all proteins will require the integration and analysis of data from a wide variety of samples.

C-HPP also aims to map specific protein variations such as post-translational modifications (PTMs), alternative splicing, and protease-processed variants.[2] Protein phosphorylation is a key regulator of cellular signal transduction processes, and its deregulation is involved in the onset and progression of various human diseases such as cancer and inflammatory and metabolic disorders.[4−7] Recent advances in proteomics, especially phosphopeptide enrichment strategies such as immobilized metal ion affinity chromatography (IMAC) and TiO_2 affinity chromatography,[8] have enabled the identification of up to several thousands of site-specific phosphorylation events within one large-scale analysis.[9−19]

As participants in C-HPP, we have integrated proteomic and phosphoproteomic analysis data from human colorectal cancer tissue and cell lines and created a chromosome-based list of identified proteins. Newly detected proteins and phosphorylated peptides were identified from the neXtProt and PhosphoSitePlus databases.

Special Issue: Chromosome-centric Human Proteome Project

Received: August 30, 2012

dx.doi.org/10.1021/pr300825v | J. Proteome Res. XXXX, XXX, XXX−XXX

MATERIALS AND METHODS

Tissue and Cell Culture Samples

Colorectal cancer tissue and tumor-adjacent normal tissue samples were obtained from 44 patients at Chiba University School of Medicine. Tissue samples were frozen in liquid nitrogen and stored at −80 °C until analysis. Written informed consent was obtained from each patient before surgery, and the protocol was approved by the ethics committees of the Proteome Research Center, National Institute of Biomedical Innovation, and the Chiba University School of Medicine. Cell cultures used were HCT116, SW480, and SW620. HCT116, a colorectal cancer cell line, was grown in RPMI 1640 medium with 10% fetal bovine serum (Invitrogen, Carlsbad, CA, USA) and penicillin/streptomycin (Invitrogen). Cells were maintained at 37 °C in an incubator supplemented with 5% CO_2 until they grew to 80% confluence. SW480 and SW620, colon cancer cell lines, were grown at 37 °C and 5% CO_2 for at least five passages in SILAC media (R1780-RPMI-1640 without arginine, lysine, leucine (Sigma−Aldrich Corp., St. Louis, MO, USA) with 10% dialyzed fetal bovine serum (Invitrogen) and 100 U/mL penicillin/streptomycin (Invitrogen)) containing 84 mg/L L-arginine (Arg0) and 40 mg/L L-lysine (Lys0) (light), or $^{13}C_6{}^{15}N_4$-L-arginine (Arg10) and $^{13}C_6$-L-lysine (Lys6) (heavy) and 50 mg/L L-leucine.

Protein Extraction and Digestion

Protein extraction and proteolytic digestion were performed using a filter-assisted sample preparation (FASP) protocol.[20] Tissue samples or pellets of cultured cells were homogenized by sonication in FASP buffer [1% SDS, 0.1 M DTT, in 0.1 M Tris/HCl, pH 7.6 and PhosSTOP phosphatase inhibitor cocktail (Roche, Mannheim, Germany)]. Protein concentration was determined using a DC protein assay kit (Bio-Rad, Richmond, CA, USA). A total of 10 mg (for phosphoproteomic analysis) or 100 μg (for proteomic analysis) of extracted proteins was digested using 1:100 (w/w) trypsin (proteomics grade; Roche) for 12 h at 37 °C. Digested peptides were concentrated and purified using a C18 Sep-PAK cartridge (Waters, Milford, MA, USA).

Phosphopeptide Enrichment

Phosphopeptide enrichment was performed using immobilized Fe(III) affinity chromatography (Fe-IMAC) as described previously.[21] The Fe-IMAC resin was prepared from Probond (Nickel-Chelating Resin; Invitrogen) by substituting Ni^{2+} on the resin with Fe^{3+}. Ni^{2+} was released from Probond upon treatment with 50 mM EDTA-2Na, and then Fe^{3+} was chelated to the ion-free resin upon incubation with 100 mM $FeCl_3$ in 0.1% acetic acid. The Fe-IMAC resin was packed into an open column for large-scale enrichment. Following equilibration of the resin with loading solution (60% acetonitrile/0.1% TFA), the peptide mixture was loaded onto the IMAC column. After washing with loading solution (9 times the volume of the IMAC resin) and 0.1% TFA (3 times the volume of the IMAC resin), phosphopeptides were eluted using 1% phosphoric acid (2 times the volume of the IMAC resin).

iTRAQ Labeling

Enriched peptides were labeled with isobaric tags for relative and absolute quantification reagents (iTRAQ 4 plex; Applied Biosystems, Foster City, CA, USA) according to the manufacturer's instructions. Peptide mixtures desalted with C18 Stage-Tips were incubated in the iTRAQ reagents for 1 h. iTRAQ 115, 116, and 117 were used for labeling individual samples, and iTRAQ 114 was used as the reference sample, a mixture of aliquots of all samples. The reaction was terminated by the addition of an equal volume of distilled water. The labeled samples were combined, acidified with trifluoroacetic acid, and desalted with C18 Stage-Tips.

Strong Cation Exchange Chromatography (SCX)

The peptides were fractionated using a HPLC system (Shimadzu Prominence UFLC) fitted with an SCX column (50 mm × 2.1 mm, 5 μm, 300 Å, ZORBAX 300SCX; Agilent Technology). The mobile phases consisted of buffer A [25% acetonitrile and 10 mM KH_2PO_4 (pH 3)] and B [25% acetonitrile, 10 mM KH_2PO_4 (pH 3), and 1 M KCl]. The labeled peptides were dissolved in 200 μL of buffer A and separated at a flow rate of 200 μL/min using a four-step linear gradient: 0% B for 30 min, 0% to 10% B in 15 min, 10% to 25% B in 10 min, 25% to 40% B in 5 min, 40% to 100% B in 5 min, and 100% B for 10 min. Fractions were collected and desalted using C18-Stage Tips (number of fractions: CRC tissues_1 peptides, 30 fractions; CRC tissues_1 phosphopeptides, 25 fractions; HCT116 peptides, 34 fractions; HCT116 phosphopeptides, 32 fractions; SW480 + SW620 peptides, 25 fractions; SW480 + SW620 phosphopeptides, 30 fractions; CRC tissues_2 non-tumor phosphopeptides, 30 fractions; CRC tissues_2 tumor phosphopeptides, 30 fractions).

LC−MS/MS Analysis

Fractionated peptides were analyzed using an LTQ-Orbitrap Velos mass spectrometer (Thermo Fisher Scientific, Bremen, Germany) equipped with a nanoLC interface (AMR, Tokyo, Japan), a nanoHPLC system (Michrom Paradigm MS2) and an HTC-PAL autosampler (CTC, Analytics, Zwingen, Switzerland). The analytical column was made in-house by packing L-column2 C18 particles (Chemical Evaluation and Research Institute (CERI), Tokyo, Japan), into a self-pulled needle (200 mm length × 100 μm inner diameter). The mobile phases consisted of buffer A (0.1% formic acid and 2% acetonitrile) and B (0.1% formic acid and 90% acetonitrile). Samples dissolved in buffer A were loaded onto a trap column (0.3 × 5 mm, L-column ODS; CERI). The nanoLC gradient was delivered at 500 nL/min and consisted of a linear gradient of buffer B developed from 5% to 30% B in 180 min. A spray voltage of 2000 V was applied.

Full MS scans were performed using an orbitrap mass analyzer (scan range m/z 350−1500, with 30 K fwhm resolution at m/z 400). The 10 most intense precursor ions were selected for the MS/MS scans, which were performed using collision-induced dissociation (CID) and higher energy collision-induced dissociation (HCD, 7500 fwhm resolution at m/z 400) for each precursor ion. The dynamic exclusion option was implemented with a repeat count of 1 and exclusion duration of 60 s. Automated gain control (AGC) was set to 1.00e + 06 for full MS, 1.00e + 04 for CID MS/MS, and 5.00e + 04 for HCD MS/MS. The normalized collision energy values were set to 35% for CID and 50% for HCD.

The CID and HCD raw spectra were extracted and searched separately against UniProtKB/Swiss-Prot (release-2010_05), which contains 20,295 sequences (the forward and reverse-decoy) of *Homo sapiens*, using Proteome Discoverer 1.3 (Thermo Fisher Scientific) and Mascot v2.3. The precursor mass tolerance was set to 7 ppm, and fragment ion mass tolerance was set to 0.5 Da for CID and 0.01 Da for HCD. The search parameters allowed two missed cleavage for trypsin,

dx.doi.org/10.1021/pr300825v | J. Proteome Res. XXXX, XXX, XXX−XXX

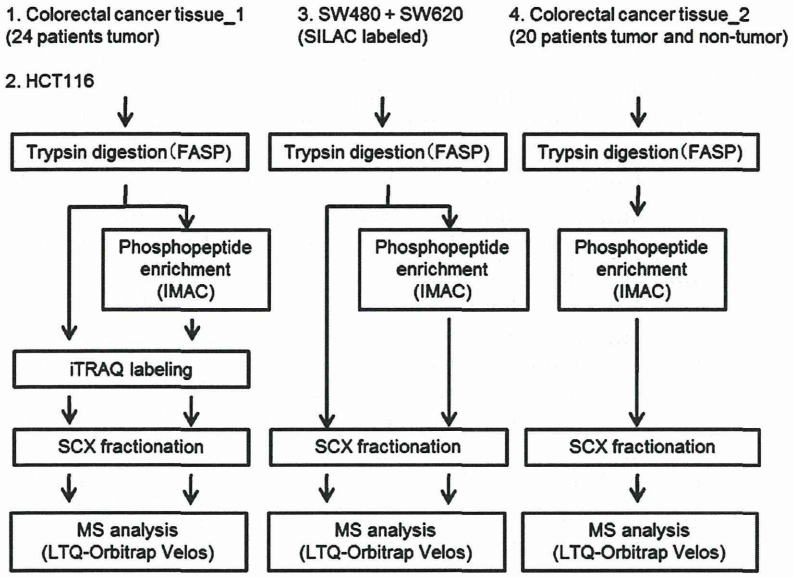


Figure 1. Schematic representation of the experimental work flow for the proteomic and phosphoproteomic analyses of the four experiments. SW480 + SW620: a mixture of protein extracts obtained from SW480 and SW620 cells. After trypsin digestion, each sample was separated for proteomic (100 μg) or phosphoproteomic (10 mg) analysis. Digested samples were separated by using an SCX column. LC–MS/MS, requiring 3-h runs, was performed using an LTQ-Orbitrap Velos.

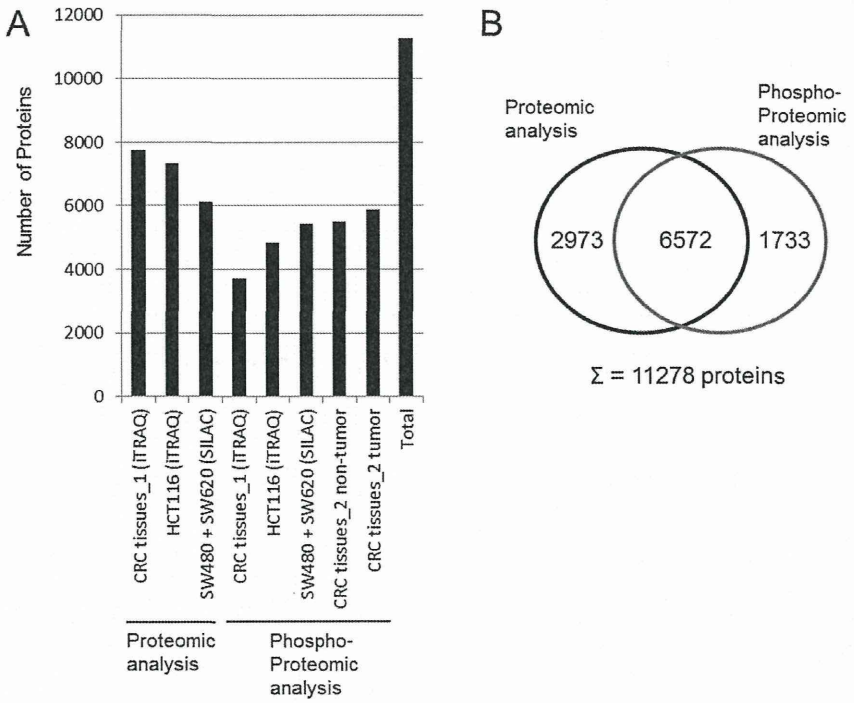


Figure 2. Number and overlap of identified proteins from the proteomic and phosphoproteomic analyses. (A) Number of identified proteins from the proteomic and phosphoproteomic analyses of the 8 data sets. (B) Proportion of proteins identified in each analysis and overlap between proteins identified by the proteomic and phosphoproteomic analyses.

fixed modifications (carbamidomethylation at cysteine), and variable modifications (oxidation at methionine). Fixed modifications were set for CRC tissue and HCT116 (iTRAQ labeling at lysine and the N-terminal residue) and SW480 + SW620 (SILAC labeling 13C(6) 15N(4) Arg, 13C(6) Lys). Variable modifications were added for phosphoproteomic analysis (phosphorylation at serine, threonine, and tyrosine). In the workflow of Proteome Discoverer 1.3, following the Mascot search, the phosphorylated sites on the identified peptides were assigned again using the PhosphoRS algorithm, which calculated the possibility of the phosphorylated site from the spectra matched to the identified peptides.[22] The score threshold for peptide identification was set at 1% false-discovery rate (FDR) and 75% phosphoRS site probability. FDR was calculated using the Percolator algorithm for peptide sequence analysis. Percolator uses >30 features of a peptide spectral match (PSM) to distinguish true positives from random matches.

dx.doi.org/10.1021/pr300825v | J. Proteome Res. XXXX, XXX, XXX–XXX

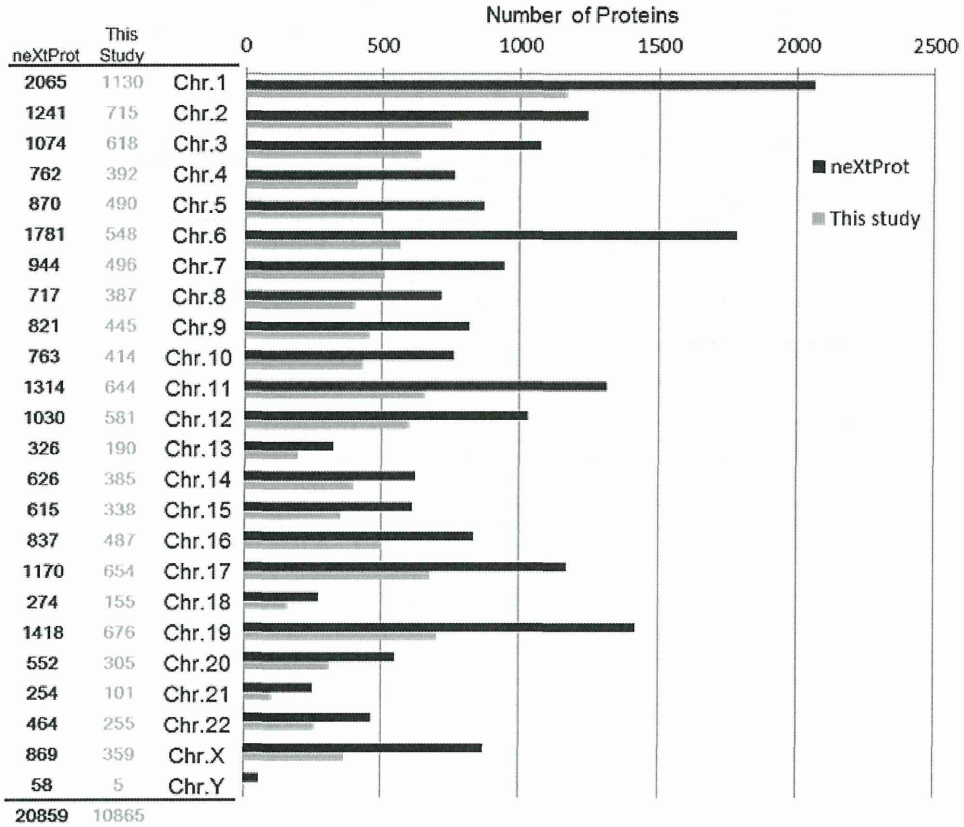


Figure 3. Chromosomal distribution of the identified proteins (gray) in relation to total proteins (black) registered in the neXtProt database.

Bioinformatics Analysis

Chromosomal locations and missing protein analyses of identified proteins were elucidated using the neXtProt database (http://www.nextprot.org/db/), and identified phosphorylation sites were elucidated using the PhosphoSitePlus database (http://www.phosphosite.org/). The function of identified missing proteins was elucidated by ingenuity pathway analysis software (Ingenuity Systems, Redwood City, USA).

Stable Isotope-Labeled Peptides

Stable isotope-labeled standard peptides (SIS peptides, crude grade) were synthesized (Thermo Fisher Scientific, Ulm, Germany). A single lysine was replaced by isotope-labeled lysine ($^{13}C_6$, 98%, $^{15}N_2$, 98%). The SIS peptides were dissolved in distilled water at a concentration of 1 $\mu g/\mu L$ and stored at −80 °C. A mixture of these SIS peptides was added to colorectal carcinoma phosphoproteomic samples.

RESULTS

As part of the C-HPP project, we combined the eight data sets from four different experiments obtained from colorectal cancer tissue and colon cancer cells; these experiments included three quantitative analyses and one non-quantitative analysis. Colorectal cancer tissues and colon cancer cells were first solubilized and trypsin-digested using the FASP method.[20] Phosphopeptides were then enriched using the IMAC method. These peptides and phosphopeptides were fractionated on a Strong Cation-Exchange (SCX) column before LC−MS/MS using an LTQ-Orbitrap mass spectrometer (Figure 1). Proteome Discoverer 1.3 software was used to analyze the RAW data files, Mascot was used as the search engine, and UniProtKB/Swiss-Prot (release-2010_05) was the database. Following data integration, 11,278 proteins were identified with Peptide FDR ≤1.0 containing at least one unique peptide corresponding to one protein in the database (Figure 2A, Supplementary Table 1−4). Of these, 8,305 proteins were identified as phosphorylated. Among the total identified proteins, 673 proteins were identified only with CID, and 386 proteins were identified only with HCD. Also, 4924 phosphopeptides were identified only with CID, and 3538 phosphopeptides were identified only with HCD. A total of 6,572 proteins were commonly identified in the proteomic and phosphoproteomic analyses (Figure 2B). However, a proportion of proteins were found not to overlap in the analyses. This is probably due to the abundance and complexity of the proteins and phosphoproteins in the samples, which prevent proteomics and phosphoproteomics to identify all of the proteins and phosphoproteins present.

Quantitative analyses were performed to investigate the differences between metastatic and non-metastatic cases by using clinical tissue and two types of cultured cells (a mixture of SW620 + SW480 and HCT116 cells). Clinical tissue samples of primary colorectal cancer obtained from 12 patients with or without metastasis were pooled. Cancers without metastasis were labeled with iTRAQ_114 or 116, and those with metastasis were labeled with iTRAQ_115 or 117. We also performed quantitative analyses between metastatic and non-metastatic cell lines. HCT116 metastatic clone was established by orthotopic implantation model mouse, and its protein expression was compared with that of the parent clone. SW620 cell line is a lymph node metastatic variant of SW480. HCT116 parent clone was labeled with iTRAQ_114 or 115, and metastatic clone was labeled with iTRAQ_116 or 117. SW480 and SW620 were reciprocally labeled with light and heavy

dx.doi.org/10.1021/pr300825v | J. Proteome Res. XXXX, XXX, XXX−XXX