Accepted manuscript: The Journal of Biochemistry 170(3) 435-443, 2021

https://doi.org/10.1093/jb/mvab081

Regular paper, biochemistry field

Unusual aggregation property of recombinantly expressed cancer-testis antigens in mammalian cells

Hannaneh Ahmadi[1], Kohei Shogen[2], Kana Fujita[2], Tomoko Honjo[1], Kazuhiro Kakimi[3], Junichiro Futami[1,2]

[1] Department of Interdisciplinary Science and Engineering in Health Systems, Okayama University, Okayama 700-8530, Japan.

[2] Department of Medical Bioengineering, Graduate School of Natural Science and Technology, Okayama University, Okayama, 700-8530, Japan.

[3]Department of Immunotherapeutics, The University of Tokyo Hospital, Bunkyo-ku, Tokyo, Japan.

Running title: Recombinant CTAs protein aggregation in mammalian cells

Corresponding author: Junichiro Futami, Ph.D.
Department of Interdisciplinary Science and Engineering in Health Systems, Okayama University, Okayama 700-8530, Japan.
Phone: 81-86-251-8217
E-mail: futamij@okayama-u.ac.jp

Abbreviations:
CLB, conventional lysis buffer; CMV, cytomegalovirus; CTA, cancer-testis antigen; EGFP, enhanced green fluorescent protein; HEK, human embryonic kidney; HKP, housekeeping proteins; HRP, horseradish peroxidase; IDP, intrinsically disordered protein; PBS, phosphate buffered saline; RT, reverse transcription

Summary

Transient expression of human intracellular proteins in human embryonic kidney (HEK) 293 cells is a reliable system for obtaining soluble proteins with biologically active conformations. Contrary to conventional concepts, we found that recombinantly expressed intracellular cancer-testis antigens (CTAs) showed frequent aggregation in HEK293 cells. Although experimental subcellular localization of recombinant CTAs displayed proper cytosolic or nuclear localization, some proteins showed aggregated particles in the cell. This aggregative property was not observed in recombinant housekeeping proteins. No significant correlation was found between the aggregative and biophysical properties, such as hydrophobicity, contents of intrinsically disordered regions, and expression levels, of CTAs. These results can be explained in terms of structural instability of CTAs, which are specifically expressed in the testis and aberrantly expressed in cancer cells and function as a hub in the protein–protein network using intrinsically disordered regions. Hence, we speculate that recombinantly expressed CTAs failed to form this protein complex. Thus, unfolded CTAs formed aggregated particles in the cell.

Keywords: aggregation, CTAs, IDPs, immunotherapy, hydrophobicity

Cancer-testis antigens (CTAs) are a heterogeneous group of proteins that are normally expressed in germline cells, typically absent or less expressed in somatic tissues, and show frequent aberrant expression in different types of cancer cells (*1, 2*). Since testicular immune privilege protects immunogenic germ cells from systemic immune attack, aberrantly expressed CTAs in cancer cells are immunogenic, eliciting both cellular and humoral immune responses in cancer patients (*3-8*). Based on this immuno-oncology significance, experimentally verified 156 types of CTAs are currently annotated in the CTdatabase (*9*). Recent cancer immunotherapy using immune checkpoint inhibitors has drastically improved clinical outcomes (*10*); however, varying response rates among individuals remain a significant issue. Biomarkers predicting clinical benefits have been investigated to evaluate individual anti-tumor immune responses. Several clinical studies have revealed that antibodies against potentially immunogenic CTAs in cancer increase along with the activation of anti-tumor immune response (*11-15*). Therefore, quantitative analysis of anti-CTA antibodies could be a useful biomarker for cancer immunity monitoring. Epitope analysis of anti-CTA antibodies in cancer patients has revealed that these antibodies recognize various linear epitopes diverse in individuals; thus, full-length CTA proteins are needed to detect polyclonal antibodies (*16*). The occurrence pattern of anti-CTA antibodies also varies among individuals. Thus, a comprehensive collection of full-length recombinant CTA proteins is required to develop an anti-CTA antibody assay system for the evaluation of cancer immunity. Preceding the preparation of an array for recombinant CTAs, we need to enhance our understanding of the biophysical properties of CTAs.

Bioinformatics analysis of CTAs to date can be summarized as follows: A total of 156 types of CTAs include 228 unique gene products. Of these, 120 (52%) have been mapped to the X chromosome (*17*), which is closely related to the loss of X-chromosome inactivation in cancer (*18*). Structural prediction of CTAs revealed that a majority of CTAs (>90%) belonged to the intrinsically disordered class. Furthermore, the CTAs are predicted to play regulatory “hub” positions in protein–protein networks (*17*).

The selection of recombinant protein expression systems can now be done from many alternatives, such as bacterial cells, mammalian cells, and cell-free expression systems. To understand mammalian intracellular proteins, the forced expression system in mammalian cells is a reliable choice. Human embryonic kidney (HEK) 293 cells are now widely employed as transient protein expression hosts because of their highly efficient transfection and robust handling in cell culture. Proper protein folding and post-translational modifications are also important in the analysis of biophysical properties of recombinant intracellular proteins under physiological conditions.

In this study, we expressed a series of recombinant intracellular CTAs in HEK293. Because of the analysis of recombinant CTAs the unexpected aggregation was observed, we tried to find out the cause of this aggregation from their protein sequence or intracellular distribution. Although the detailed aggregation property failed to find only from the protein sequence of CTAs, this information will help in understanding the immunological response of CTAs as well as the preparation of recombinant CTA protein arrays for the analysis of anti-CTA antibody biomarker detections (19).

Materials and Methods

Recombinant expression of CTAs and housekeeping proteins

The cDNAs encoding human and mouse CTAs, human housekeeping proteins (HKPs) (Table I), and enhanced green fluorescent protein (EGFP, Uniprot: C8CHS1) were amplified by PCR from normal human testis first-strand cDNA (BioChain, CA, USA) or normal mouse testis first-strand cDNA (GenoStaff, Tokyo, Japan). Recombinant proteins were designed to possess His-tag (MGSSHHHHHHSSGLVPRGSH) at the amino-terminus and StrepTagII (GPGWSHPQFEK) on the carboxyl-terminus. DNA was cloned into the previously developed super gene expression vector (*20, 21*), with the cytomegalovirus (CMV) promoter developed by inserting the triple translational enhancer sequences of hTERT, SV40, and CMV downstream of the sequence of the bovine growth hormone polyA gene. Recombinant proteins were transiently expressed in HEK293 cells (FreeStyle 293F, Life Technologies, Carlsbad, CA, USA) using FreeStyle F17 Medium and 293fectin transfection reagent (Life Technologies) according to the manufacturer's instructions. After the transfection of expression plasmid DNA in 2 mL cell culture in 6-well plates, the cells were cultivated using an orbital shaker (125 rpm) at 37°C in the presence of 8% CO_2 for 2 days. In order to quantify the recombinant protein expression levels of inter-experiments, EGFP cloned in the same vector was employed for normalization. Harvested cells were then dissolved in conventional lysis buffer (CLB, 20 mM HEPES buffer, pH 7.5, 500 mM NaCl, 1% NP-40) supplemented with protease inhibitor cocktail (Nacalai Tesque, Kyoto, Japan) and subsequently disrupted on ice using a sonicator (Bioruptor UCD-250, CosmoBio, Tokyo, Japan) to reduce viscosity from genomic DNA. The lysates were fractionated into soluble and insoluble fractions by centrifugation at 14,500 × *g* for 10 min at 4°C [Fig. 1(A)].

Evaluation of the effect of lysis buffer on the solubility of CTAs

The effect of lysis buffer on the solubility of recombinant CTAs was confirmed using three recombinant human CTAs, CT40, CT52, and CT55 (Table I), and GFP as a control. HEK293 cells expressing these proteins were dissolved in four typical lysis buffers: phosphate buffered saline (PBS), CLB, radioimmunoprecipitation assay buffer (RIPA, 50 mM Tris-HCl buffer, pH 7.0, 150 mM NaCl, 1% Nonidet® P40, 0.5% sodium deoxycholate, and 0.1% SDS), or M-PER Mammalian Protein Extraction Reagent (Thermo Scientific, MA, USA) supplemented with a protease inhibitor cocktail (Nacalai Tesque). The cells were subsequently disrupted on ice using a sonicator (Bioruptor UCD-250, Cosmo Bio). The lysates were fractionated into soluble and precipitated fractions by centrifugation at 14,500 × *g* for 10 min at 4°C. The samples were then separated by SDS-PAGE under reducing conditions. The effect of each lysis buffer on the solubility of recombinant proteins in cell lysates was confirmed by western blotting using anti-His-tag antibody (OGHis, MBL, Nagoya, Japan).

Quantification of intracellular solubility of recombinant proteins and biophysical predictions

The ratio of soluble and precipitated fractions of recombinant CTAs or HKPs in HEK293 cells was analyzed by western blotting after lysing cells in CLB, as described above. The target band reacted by anti-His-tag antibody was illuminated by anti-mouse IgG, horseradish peroxidase (HRP)-linked antibody (Cell Signaling Technology, MA, USA), and chemiluminescence reagent (Western Lightning Plus-ECL; PerkinElmer, MA, USA), and detected using a luminescent image analyzer (LAS-4000 mini; FUJIFILM, Tokyo, Japan). The ratios of soluble and precipitated fractions of each recombinant protein were determined using the NIH Image J software (http://imagej.nih.gov/ij/), by following the ImageJ user guide in the case of Gels (part 30.13). The recombinant protein expression levels were evaluated by combining band intensities from soluble and precipitated fractions, and then normalized inter-experimental variations by the reference band intensities from His-tag fused EGFP transfected cell lysates. The evaluated band intensities were classified as high (>5 × 10^7AU), mid (2× 10^7–5 × 10^7AU), or low (<2 × 10^7 AU) levels using a Multi Gauge-Ver3.0 (Fujifilm, Tokyo, Japan).

Recombinant protein disordered regions were predicted by two different algorithms, namely the FoldIndex (*22*) and Regional Order Neural Network (*23*). The range of full-length protein hydrophobicity was determined by the grand average of hydropathy value for protein sequences (http://www.gravy-calculator.de/index.php), calculated by adding the hydropathy value for each residue and dividing by the length of the sequence (*24*). The amount of theoretical PI, Cys

residues, and protein length of each protein were computed by using the Expasy tool (https://www.expasy.org/). These data are provided in Supplementary data 1.

Analysis of subcellular localization of recombinant CTAs

Fourteen human CTAs with diverse solubility (CT1.6, CT96, CT11, CT26, CT8, CT76, CT86, CT53, CT9, CT57, CT79, CT106, CT33, and CT55) were transiently transfected in HeLa (HeLa S3, ATCC) cells on glass-based dish using Lipofectamine 3000 transfection reagent (Invitrogen, MA, USA) according to the manufacturer's instructions. After culturing for 48 h after transfection, cells were fixed with 4% paraformaldehyde phosphate buffer solution (Wako, Osaka, Japan) and then permeabilized with 0.1% Triton X-100 in PBS for 30 min at room temperature. Cells were blocked with 1% BSA in PBS at 4°C overnight, and intracellular recombinant CTAs were detected using 1 μg/mL of anti-His-tag mAb-Alexa Fluor® 488 (clone OGHis, MBL) and then visualized by confocal laser-scanning microscopy LSM510 META (Carl Zeiss, Jena, Germany).

Analysis of intracellular solubility of endogenous CTAs

The candidates for endogenous expression of CTAs in HeLa or HEK293T (PEAKrapid, ATCC) cells were selected from mRNA expression in cell lines (RT-PCR) data in CTdatabase (http://www.cta.lncc.br/) or mRNA expression overview in The Human Protein Atlas database (https://www.proteinatlas.org/), and mRNA expression was confirmed by reverse transcription (RT)-PCR. Total RNA was extracted from exponentially growing cells using the RNeasy Plus Mini Kit (Qiagen, Hilden, Germany). First-strand cDNA was synthesized using the ReverTra Ace® qPCR RT Master Kit (TOYOBO, Osaka, Japan). Primers used to amplify each cDNA are listed in Supplementary data 2. Rabbit polyclonal antibodies to CTAs were prepared at Cosomo Bio (Tokyo, Japan) by immunization with full-length antigens purified by previously described procedures (*19*). The solubility of endogenous CTAs in the cell lysates was determined under the same conditions for recombinant CTA analysis, except for the use of specific polyclonal antibody, anti-rabbit IgG, and HRP-linked antibody (Cell Signaling Technology).

Results

Recombinant expression of CTAs in HEK293 cells and analysis of their intracellular solubility

Transiently expressed recombinant proteins in HEK293 cells with approximately 70% transfection efficiency observed on EGFP-transfected cells were analyzed by western blotting. To evaluate the effect of lysis buffer selection on protein solubility, four typical lysis buffers were compared. As shown in Fig. 1B, the ratio of recombinant proteins in soluble or precipitated fractions was detected at levels comparable in all the examined cases. The control experiment of EGFP showed partially precipitated fractions due to the tendency of many fluorescent proteins

to oligomerize, which can result in aggregation (*25*). Therefore, CLBs prepared in house were used in subsequent experiments.

Unusual intracellular solubility of recombinant CTAs

A series of 70 human CTAs annotated as cytoplasmic or nuclear proteins (Table I) were transiently expressed in HEK293 cells for 2 days. This condition was chosen as the highest intracellular recombinant protein expression with minimum cell die and protein degradations. After lysing the cells in CLB by sonication, the lysates were separated into soluble and precipitated fractions. All human CTAs showed sufficient expression, as detected by western blotting, and showed a diverse range of intracellular solubility, from 17.4% to 86.5% (Fig. 2A). These solubility trends were similar for the 15 types of mouse-derived CTAs (Fig. 2B). In contrast, 12 human intracellular HKPs expressed in the same expression vectors and protein expression and extraction conditions, showed high solubility without any detectable precipitates (Fig. 2C). Soluble recombinant proteins on this condition confirmed to pass through the 0.22 μm filter, but a part of proteins was trapped on the ultrafiltration membrane (MWCO 100 kDa). These suggesting that these soluble proteins in lysate are monomeric to oligomeric dispersion states (Supplemental Fig.1). Taken together, these results show that CTAs show unusual aggregation-favored properties on recombinant protein expression in HEK293 cells (Fig. 3A).

Relationship between the intracellular solubility and physicochemical properties of recombinant CTAs

To determine the physicochemical relevance of the intracellular solubility of recombinant CTAs, we analyzed the contents of predicted disordered regions (Fig. 3B) and hydrophobicity of polypeptide chains (Fig. 3C). Another typical physicochemical index including theoretical isoelectric point, Cys contents in protein sequence, and polypeptide chain length failed to find a correlation between their solubility (Supplemental Fig.2). As predicted, the majority of CTAs had intrinsically disordered regions, and the amount of disorder in CTAs was relatively higher than that in HKPs. However, the diversity in the disordered regions of CTAs had no correlation with their intracellular solubility. Hydrophobicity of polypeptides was also observed; HKPs with ordered structures possessed relatively higher hydrophobicity than CTAs, but no correlation was found between hydrophobicity and intracellular solubility of CTAs. Overexpression of recombinant proteins also causes aggregation of intracellular proteins. Although recombinant protein expression driven by the CMV promoter should be categorized as overexpression, the

protein expression levels can be categorized into three levels. However, no correlation was found between the protein expression level and solubility (Fig. 3D).

Proper Intracellular distribution of recombinant CTAs with aggregated particles

The intracellular distribution of recombinant CTAs expressed in HeLa cells was analyzed using immunofluorescence microscopy (Fig. 4A). On analysis of the 14 selected CTAs from higher solubility (86.5%) to lower solubility (19.6%), the majority of CTAs showed proper intracellular distributions matching the endogenous protein subcellular localizations annotated in UniProt (https://www.uniprot.org/) (Table I). Nuclear proteins CT53, CT9, and CT33 showed nuclear localization with vesicle-like particles in the nucleus. Cytoplasmic proteins such as CT57 and CT79 showed a cytosolic distribution with visible aggregated particles. CT96, CT8, CT26, and CT106 showed partial nuclear localization and mainly cytosolic distribution with visible aggregated particles. CT11, CT76, and CT86 showed partial cytosolic localization and mainly nuclear distribution with a dot-like structure. On the other hand, the intracellular distribution of two human intracellular HKPs, SRM and CSTB, showed both cytosolic and nuclear localization with high solubility and no detectable precipitates (Fig. 4B). Taken together, these results indicate that some recombinant CTAs showed intracellular aggregation, but this was not observed for HKPs. Thus, the unusual intracellular solubility of CTAs is a characteristic property of this family of proteins.

Unusually high intracellular solubility of endogenous CTAs

A series of six human CTAs endogenously expressed in HeLa and HEK293 cells were selected from The Human Protein Atlas database, and mRNA expression was confirmed by RT-PCR (Fig. 4C). After lysing the cells in CLB by sonication, the lysates were separated into soluble and precipitated fractions. All six CTAs showed sufficient expression clearly detected by western blotting using rabbit polyclonal antibodies, except for CT53 in HEK293 cells due to lower expression the other CTAs. The endogenously expressed CTAs showed higher intracellular solubility than the recombinantly expressed CTAs in HEK293 (Fig. 4C). Importantly, endogenous CTAs in HeLa cells showed detectable P fractions, strongly suggesting that aberrantly expressed CTAs in the cancer cell line also tended to aggregate in cells. The CMV-promoter derived acute expression was suggested to be independent of the cause of aggregation because the endogenous expression level of CT57 in HeLa cell was 1.5 % of recombinant protein, but their solubility level is similar (Fig 4C and Supplemental Fig.3).

Discussion

In this study, we discovered that recombinantly expressed intracellular CTAs show frequent aggregation in HEK293 cells. We performed quantitative analysis by western blotting of cell lysates and observed intracellular distributions with aggregated particles in cells by microscopic immunofluorescence analysis. This aggregation-favored property is typical of CTAs, regardless of their predicted intrinsically disordered protein (IDP) region contents, hydrophobicity, and expression level. It is generally accepted that overexpression of functional recombinant mammalian membrane proteins has poor chances of success, and aggregation and misfolding are often observed due to the hydrophobic nature of transmembrane segments (*26, 27*). Therefore, we excluded CTAs possessing transmembrane regions from this study (Table I). Overexpression of recombinant protein in mammalian cells using a strong CMV promoter system or tagging of peptide for detection could be a cause of intracellular aggregation. However, we confirmed that the recombinant expression system in this study showed no aggregation in 12 types of human intracellular HKPs, but aggregation was frequently observed in intracellular CTAs.

We speculate that the structural instability of CTAs is the main cause of the frequent aggregation of recombinant CTAs in cells. Many testis-specific proteins work cooperatively during spermatogenesis. As predicted before, CTAs originally work as regulatory “hub” positions in protein-protein networks (*17*). This means that recombinant CTAs and aberrantly expressed CTAs in cancer cells potentially lack binding partners in the cell and form aggregation *in vivo*. Some endogenously and aberrantly expressed CTAs, such as CT53 and CT57, showed similar low solubility on their recombinant expression (Fig.4C). Conversely, endogenously expressed CTAs, such as CT86, CT102, CT106, and CT109, showed higher solubility compared to their recombinant expression (Fig.4C). This controversial result may be explained by the synergistic effect of structurally unstable CTAs and an acute recombinant protein expression, resulting in intracellular aggregation of CTAs.

In mammalian cells, misfolded proteins are considered to be toxic (*28-30*). Therefore, mammalian cells possess degradation pathways for the removal of misfolded proteins, including the ubiquitin-proteasome system, chaperone-mediated autophagy, and macroautophagy (*31*). The aggregation of misfolded proteins is also employed as protection system cells due to misfolded protein toxicity (*32, 33*). It has been reported that proteins with somatic mutations often cause intracellular aggregation that can occur in a diseased state or during normal physiological functions. This mechanism is clearly related to the folding ability of proteins because the wild type counterpart no longer shows aggregation (*31, 34*). Taken together these mechanisms on misfolded protein aggregates in mammalian cells, recombinant CTAs aggregated

particles observed in the cells are categorized to the compartment of insoluble protein deposit (IPOD) (*32, 33*). It is unclear the fate of IPOD in mammalian cells but aberrantly expressed CTAs have the potential to form IPOD in the cancer cells. CTAs in cancer cells are known to show humoral immune responses in cancer patients. Although the less expression of CTAs in normal tissues is the main reason for their immunogenicity, the aggregation-favored property of CTAs may involve their immunogenicity because aggregated proteins show immunogenicity (*35*). Such experiments are essential for understanding the biophysical properties of recombinant CTAs, as well as the immuno-oncology aspects of CTAs contributed by protein properties.

Reference

(1) Scanlan, M.J., Simpson, A.J., and Old, L.J. (2004) The cancer/testis genes: review, standardization, and commentary. *Cancer Immun*. **4**, 1

(2) Simpson, A.J., Caballero, O.L., Jungbluth, A., Chen, Y.T., and Old, L.J. (2005) Cancer/testis antigens, gametogenesis and cancer. *Nat Rev Cancer*. **5**, 615-625

(3) Chen, Y.T., Güre, A.O., Tsang, S., Stockert, E., Jäger, E., Knuth, A., and Old, L.J. (1998) Identification of multiple cancer/testis antigens by allogeneic antibody screening of a melanoma cell line library. *Proc Natl Acad Sci U S A*. **95**, 6919-6923

(4) Caballero, O.L., and Chen, Y.T. (2009) Cancer/testis (CT) antigens: potential targets for immunotherapy. *Cancer Sci*. **100**, 2014-2021

(5) Krishnadas, D.K., Bai, F., and Lucas, K.G. (2013) Cancer testis antigen and immunotherapy. *Immunotargets Ther*. **2**, 11-19

(6) Chen, H., Werner, S., Tao, S., Zörnig, I., and Brenner, H. (2014) Blood autoantibodies against tumor-associated antigens as biomarkers in early detection of colorectal cancer. *Cancer Lett*. **346**, 178-187

(7) Gjerstorff, M.F., Andersen, M.H., and Ditzel, H.J. (2015) Oncogenic cancer/testis antigens: prime candidates for immunotherapy. *Oncotarget*. **6**, 15772-15787

(8) Wei, X., Chen, F., Xin, K., Wang, Q., Yu, L., Liu, B., and Liu, Q. (2019) Cancer-Testis Antigen Peptide Vaccine for Cancer Immunotherapy: Progress and Prospects. *Transl Oncol*. **12**, 733-738

(9) Almeida, L.G., Sakabe, N.J., deOliveira, A.R., Silva, M.C., Mundstein, A.S., Cohen, T., Chen, Y.T., Chua, R., Gurung, S., Gnjatic, S., Jungbluth, A.A., Caballero, O.L., Bairoch, A., Kiesler, E., White, S.L., Simpson, A.J., Old, L.J., Camargo, A.A., and Vasconcelos, A.T. (2009) CTdatabase: a knowledge-base of high-throughput and curated data on cancer-testis antigens. *Nucleic Acids Res*. **37**, D816-819

(10) Brahmer, J., Reckamp, K.L., Baas, P., Crinò, L., Eberhardt, W.E., Poddubskaya, E., Antonia, S., Pluzanski, A., Vokes, E.E., Holgado, E., Waterhouse, D., Ready, N., Gainor, J., Arén Frontera, O., Havel, L., Steins, M., Garassino, M.C., Aerts, J.G., Domine, M., Paz-Ares, L., Reck, M., Baudelet, C., Harbison, C.T., Lestini, B., and Spigel, D.R. (2015) Nivolumab versus Docetaxel in Advanced Squamous-Cell Non-Small-Cell Lung Cancer. *N Engl J Med*. **373**, 123-135

(11) Gunawardana, C.G., Memari, N., and Diamandis, E.P. (2009) Identifying novel autoantibody signatures in ovarian cancer using high-density protein microarrays. *Clin Biochem*. **42**, 426-429

(12) Gnjatic, S., Ritter, E., Büchler, M.W., Giese, N.A., Brors, B., Frei, C., Murray, A., Halama, N., Zörnig, I., Chen, Y.T., Andrews, C., Ritter, G., Old, L.J., Odunsi, K., and Jäger, D. (2010) Seromic profiling of ovarian and pancreatic cancer. *Proc Natl Acad Sci U S A*. **107**, 5088-5093

(13) Antony, F., Deantonio, C., Cotella, D., Soluri, M.F., Tarasiuk, O., Raspagliesi, F., Adorni, F., Piazza, S., Ciani, Y., Santoro, C., Macor, P., Mezzanzanica, D., and Sblattero, D. (2019) High-throughput assessment of the antibody profile in ovarian cancer ascitic fluids. *Oncoimmunology*. **8**, e1614856

(14) Ghassem-Zadeh, S., Hufnagel, K., Bauer, A., Frossard, J.L., Yoshida, M., Kutsumi, H., Acha-Orbea, H., Neulinger-Muñoz, M., Vey, J., Eckert, C., Strobel, O., Hoheisel, J.D., and Felix, K. (2020) Novel Autoantibody Signatures in Sera of Patients with Pancreatic Cancer, Chronic Pancreatitis and Autoimmune Pancreatitis: A Protein Microarray Profiling Approach. *Int J Mol Sci*. **21**,

(15) Ohue, Y., Kurose, K., Karasaki, T., Isobe, M., Yamaoka, T., Futami, J., Irei, I., Masuda, T., Fukuda, M., Kinoshita, A., Matsushita, H., Shimizu, K., Nakata, M., Hattori, N., Yamaguchi, H., Fukuda, M., Nozawa, R., Kakimi, K., and Oka, M. (2019) Serum Antibody Against NY-ESO-1 and XAGE1 Antigens Potentially Predicts Clinical Responses to Anti-Programmed Cell Death-1 Therapy in NSCLC. *J Thorac Oncol*. **14**, 2071-2083

(16) Kawabata, R., Wada, H., Isobe, M., Saika, T., Sato, S., Uenaka, A., Miyata, H., Yasuda, T., Doki, Y., Noguchi, Y., Kumon, H., Tsuji, K., Iwatsuki, K., Shiku, H., Ritter, G., Murphy, R., Hoffman, E., Old, L.J., Monden, M., and Nakayama, E. (2007) Antibody response against NY-ESO-1 in CHP-NY-ESO-1 vaccinated patients. *Int J Cancer*. **120**, 2178-2184

(17) Rajagopalan, K., Mooney, S.M., Parekh, N., Getzenberg, R.H., and Kulkarni, P. (2011) A majority of the cancer/testis antigens are intrinsically disordered proteins. *J Cell Biochem*. **112**, 3256-3267

(18) Kang, J., Lee, H.J., Jun, S.Y., Park, E.S., and Maeng, L.S. (2015) Cancer-Testis Antigen Expression in Serous Endometrial Cancer with Loss of X Chromosome Inactivation. *PLoS One*. **10**, e0137476

(19) Futami, J., Nonomura, H., Kido, M., Niidoi, N., Fujieda, N., Hosoi, A., Fujita, K., Mandai, K., Atago, Y., Kinoshita, R., Honjo, T., Matsushita, H., Uenaka, A., Nakayama, E., and Kakimi, K. (2015) Sensitive Multiplexed Quantitative Analysis of Autoantibodies to Cancer Antigens with Chemically S-Cationized Full-Length and Water-Soluble Denatured Proteins. *Bioconjug Chem*. **26**, 2076-2084

(20) Sakaguchi, M., Watanabe, M., Kinoshita, R., Kaku, H., Ueki, H., Futami, J., Murata, H., Inoue, Y., Li, S.A., Huang, P., Putranto, E.W., Ruma, I.M., Nasu, Y., Kumon, H., and Huh, N.H. (2014) Dramatic increase in expression of a transgene by insertion of promoters downstream of the cargo gene. *Mol Biotechnol.* **56**, 621-630

(21) Kinoshita, R., Watanabe, M., Huang, P., Li, S.A., Sakaguchi, M., Kumon, H., and Futami, J. (2015) The cysteine-rich core domain of REIC/Dkk-3 is critical for its effect on monocyte differentiation and tumor regression. *Oncol Rep.* **33**, 2908-2914

(22) Prilusky, J., Felder, C.E., Zeev-Ben-Mordehai, T., Rydberg, E.H., Man, O., Beckmann, J.S., Silman, I., and Sussman, J.L. (2005) FoldIndex: a simple tool to predict whether a given protein sequence is intrinsically unfolded. *Bioinformatics.* **21**, 3435-3438

(23) Yang, Z.R., Thomson, R., McNeil, P., and Esnouf, R.M. (2005) RONN: the bio-basis function neural network technique applied to the detection of natively disordered regions in proteins. *Bioinformatics.* **21**, 3369-3376

(24) Kyte, J., and Doolittle, R.F. (1982) A simple method for displaying the hydropathic character of a protein. *J Mol Biol.* **157**, 105-132

(25) Chudakov, D.M., Matz, M.V., Lukyanov, S., and Lukyanov, K.A. (2010) Fluorescent proteins and their applications in imaging living cells and tissues. *Physiol Rev.* **90**, 1103-1163

(26) Andréll, J., and Tate, C.G. (2013) Overexpression of membrane proteins in mammalian cells for structural studies. *Mol Membr Biol.* **30**, 52-63

(27) Grisshammer, R. (2006) Understanding recombinant expression of membrane proteins. *Curr Opin Biotechnol.* **17**, 337-340

(28) Shigemitsu, Y., and Hiroaki, H. (2018) Common molecular pathogenesis of disease-related intrinsically disordered proteins revealed by NMR analysis. *J Biochem.* **163**, 11-18

(29) Takeuchi, T. (2020) Pathogenic and protective roles of extracellular vesicles in neurodegenerative diseases. *J Biochem.*

(30) Treusch, S., Cyr, D.M., and Lindquist, S. (2009) Amyloid deposits: protection against toxic protein species? *Cell Cycle.* **8**, 1668-1674

(31) Migas, U.M., Quinn, M.K., and McManus, J.J. (2017) Protein self-assembly following in situ expression in artificial and mammalian cells. *Integr Biol (Camb).* **9**, 444-450

(32) Kaganovich, D., Kopito, R., and Frydman, J. (2008) Misfolded proteins partition between two distinct quality control compartments. *Nature.* **454**, 1088-1095

(33) Tyedmers, J., Mogk, A., and Bukau, B. (2010) Cellular strategies for controlling protein aggregation. *Nat Rev Mol Cell Biol.* **11**, 777-788

(34) Ciechanover, A., and Kwon, Y.T. (2015) Degradation of misfolded proteins in neurodegenerative diseases: therapeutic targets and strategies. *Exp Mol Med*. **47**, e147

(35) Moussa, E.M., Panchal, J.P., Moorthy, B.S., Blum, J.S., Joubert, M.K., Narhi, L.O., and Topp, E.M. (2016) Immunogenicity of Therapeutic Protein Aggregates. *J Pharm Sci*. **105**, 417-430

Figure Legends

Fig. 1. Recombinant expression of CTAs and HKPs in HEK293 cells

A) Schematic procedure of transient expression of CTAs in HEK293 cells

B) Western blotting analysis of three CTAs that were transiently expressed in HEK293 cells and lysed in four different kinds of lysis buffer using anti-His-tag antibody. GFP was used as a control.

Fig. 2. Intracellular solubility of recombinant CTAs and HKPs in HEK293 cells

Fig. 3. Relationship between intracellular solubility and physicochemical properties of recombinant CTAs

A) Intercellular solubility percentage of CTAs and HKPs transiently expressed in HEK293 cells compared with (B) predicted disordered regions, (C) range of hydrophobicity, and (D) their expression levels.

Fig. 4. Analysis of intracellular distribution and solubility of recombinant and endogenous CTAs.

Immunofluorescent staining for (A) human CTAs (B) and two HKPs transiently transfected in HeLa cells and detected using anti-His-tag mAb-Alexa Fluor® 488 antibody (Scale bar: 10 μm).

C) Western blotting analysis of six human CTAs endogenously expressed in HeLa and HEK293 cells and lysed in CLB using rabbit polyclonal antibodies. mRNA expression of CTAs in HeLa and HEK293 cells was confirmed by RT-PCR (Right panel pasted from Fig. 2A). ND means not detectable.

Acknowledgements

This work is partially supported by JST START Grant Number JPMJST1918 and JSPS KAKENHI (Grant–in-Aid for Scientific Research (B), No. 16H04580 (J. Futami).

Author contributions

H.A. K.K. and J.F. conceived the project. H.A., K.S., K.F., and T.H performed experiments. H.A. and J.F. wrote the manuscript. All authors read and approved the final manuscript.

Competing interests

The authors declare no competing financial interests.

Tables

Table 1: Human CTAs, Mouse CTAs and HKPs that have been expressed for solubility checking

Human CTAs number	Solu0bility (%)	Uniprot ID	MW	Subcellular location (Uniprot)
CT1.6	86.5	P43360	34,891 Da	Unknown
CT3.1	68.2	P43366	39,037 Da	Unknown
CT5.4	54.5	Q16385	21,620 Da	Nucleus
CT8	48.4	Q15431	114,192 Da	Nucleus - chromosome
CT9	38.5	Q58F21	107,954 Da	Nucleus
CT11	53.3	Q9BXN6	11,029 Da	Nucleus -Cytoplasm
CT12.2	70.1	Q96GT9	12,354 Da	Unknown
CT13	77.8	Q9NXZ2	72,844 Da	Unknown
CT26	52.2	Q86TM3	71,154 Da	Nucleus-Cytosol
CT28	29.7	Q9P127	35,937 Da	Nucleus - Cytoplasm
CT30	38.6	Q5H9I0	44,967 Da	Nucleus - Cytoplasm
CT32	32	P07864	36,311 Da	Cytoplasm
CT33	24.7	Q86VD1	112,881 Da	Nucleus
CT35	40	Q9Y5K1	44,537 Da	Nucleus
CT38	18.3	Q9BXU8	21,142 Da	Cytoplasm
CT39	64.4	Q9GZY0	71,627 Da	Nucleus - Cytoplasm
CT40	50.1	Q5H9L4	52,588 Da	Nucleus -Cytoplasm
CT47	77.2	Q5JQC4	30,100 Da	Unknown
CT51	37.8	Q68G75	17,486 Da	Nucleus - Cytoplasm
CT52	27.2	Q8TBZ0	96,726 Da	Nucleus

CT53	39.4	P49910	55,771 Da	Nucleus
CT55	19.6	Q8WUE5	29,052 Da	Unknown
CT56	71.3	Q9P2T0	43,444 Da	Nucleus
CT57	31	Q9H568	41,360 Da	Cytoskeleton-Cytoplasm
CT59	46.1	Q6YFQ2	10,529 Da	Mitochondria -intermembrane space
CT63	66.9	Q8IV76	87,428 Da	Nucleus
CT65	35.2	O00295	58,664 Da	Cytoplasm-secreted
CT68	42.6	Q8TBY0	60,023 Da	Nucleus
CT72	28.1	Q9BXA6	30,331 Da	Nucleus
CT74	30.6	Q8IYA8	66,343 Da	Chromosome
CT76	40.7	Q8N0S2-2	39,699 Da	Nucleus - Chromosome
CT77	50.9	Q8N123	34,727 Da	Unknown
CT78	69.9	Q01534	35,012 Da	Nucleus - Cytoplasm
CT79	28.3	Q9BZW7	81,421 Da	Cytoskeleton - Cytoplasm
CT80	38.6	Q8TC59	109,849 Da	Cytoplasm-Nucleus
CT81	69.9	Q5W041	96,405 Da	Extra cellular region
CT82	60.8	O75969	94,751 Da	Nucleus
CT84	71.5	Q96KB5	36,085 Da	Nucleus
CT86	40.5	Q5H943	24,691 Da	Nucleus - Chromosome
CT88	70	O75952	52,774 Da	Nucleus- Cytoskeleton
CT91	49.7	Q9HAT0	23,893 Da	Nucleus-Cytoplasm
CT94.1	48	P04553	6,823 Da	Nucleus-Chromosome
CT94.2	45.6	P04554	13,051 Da	Nucleus-Chromosome

CT96	70.9	P33981	97,072 Da	Cytoskeleton -Nucleus
CT100	72.2	Q7Z7J5	33,784 Da	Nucleus
CT102	47	Q5T6F0	50,517 Da	Cytoskeleton -Cytoplasm
CT104.2	57.9	Q6S8J3-3	41,909 Da	Extra cellular region
CT106	27.3	Q9BZD4	54,304 Da	Nucleus-Cytosol
CT107	72.3	Q9BQY4	31,692 Da	Nucleus
CT109	17.4	Q6P9F0	77,748 Da	Nucleus-Cytoplasm
CT110	43.4	Q9NW75	58,944 Da	Nucleus
CT112	46.2	Q8NEK8	44,500 Da	Unknown
CT113	30.5	Q8IWB6	16,901 Da	Cytoplasm
CT115	45.3	Q8N9E0	28,941 Da	Unknown
CT117	67.6	Q5TZF3-2	29,777 Da	Unknown
CT129	60.6	Q92600	33,631 Da	Nucleus
CT130	32.4	P78395	57,890 Da	Nucleus
CT132	28.9	Q7Z5L4	19,186 Da	Mitochondrion
CT133	26.0	Q14990	28,366 Da	Nucleus- Cytoskeleton
CT135	60.5	Q96PU9	27,710 Da	Cytoplasm
CT138	37.3	Q8N7E2	48,875 Da	Cytoplasm
CT139	53.1	Q99661	81,313 Da	Nucleus- Cytoskeleton
CT141	49.6	O75602	55,476 Da	Cytoskeleton
CT144	46.3	Q8N4B4	52,646 Da	Unknown
CT148	35.8	Q8IWF9	48,851 Da	Unknown
CT149	38.5	Q96M29	56,294 Da	Nucleus-Flagellum

CT152	62.6	P59282	18,503. Da	Cytosol
CT154	49.3	Q9Y5R6	39.473 Da	Nucleus
CT155	61.8	Q9BXL5	55,341 Da	Nucleus
CT156	54.4	Q8NHS0	25,686 Da	Cytosol-Nucleus

Mouse CTAs number	**Solubility (%)**	**Uniprot ID**	**MW**	**Subcellular location**
CT1.3	79.29	Q6T340	36,207Da	Unknown
CT1.4	73.02	Q3SXV1	35,677Da	Nucleus -Cytosol
CT3.5	55.8	Q9D2H4	35,851Da	Unknown
CT8	67.35	Q62209	115,935Da	Nucleus-Chromosome
CT79	29.37	Q6NY15-3	72,295Da	Cytoplasm-Cytoskeleton
CT5.4	38.45	Q8C5Z3	19,584Da	Nucleus
CT22	96.35	Q62252	17,296Da	Nucleus-Chromosome
CT33	37.51	Q9WVL5	108,310Da	Nucleus
CT38	42.84	Q99MX2	20,491Da	Cytoplasm
CT46	42.64	Q9D5T7	44,933Da	Nucleus
CT51	57.36	Q9DAM3	18,493Da	Nucleus-Cytoplasm
CT88	81.51	Q9D424	48,333Da	Cytoskeleton-Cytoplasm
CT106	32.96	Q99P69	54,594Da	Nucleus
CT128	59.83	Q8BVN9	49,371Da	Nucleus-Cytoplasm

CT149	39.61	G5E8A8	62,734Da	Unknown

HK Proteins	**Solubility (%)**	**Uniprot ID**	**MW**	**Subcellular location**
GAPDH	100	P04406	36,053 Da	Nucleus - Cytosol
ACTB	100	P60709	39,037 Da	Nucleus - Cytoskeleton
RPL19	100	P84098	23,466 Da	Cytosol-Nucleus
LDHB	100	P07195	36,638 Da	Cytosol-Nucleus
SRRM1	100	Q8IYB3	102,335 Da	Nucleus-Cytosol
LDHA	100	P00338	36,689 Da	Cytoplasm
CSTB	100	P04080	11,140 Da	Nucleus-Cytosol
RPS27A	100	P62979	17,965 Da	Nucleus - Cytoplasm
ALDOA	100	P04075	39,420 Da	Cytosol-Nucleus
ODC1	100	P11926	51,148 Da	Cytosol
PGK1	100	P00558	44,615 Da	Cytoplasm
CKB	100	P12277	42,644 Da	Cytoplasm

Figure1

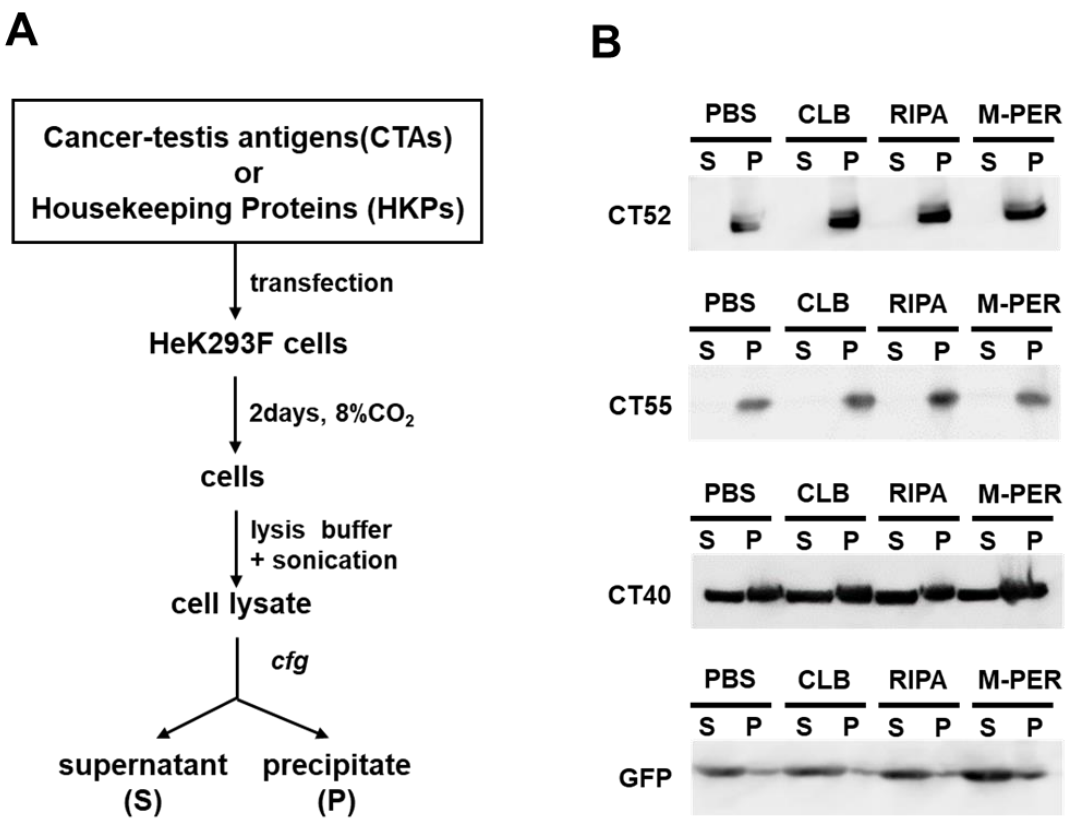

A
Cancer-testis antigens(CTAs)
or
Housekeeping Proteins (HKPs)
transfection
HeK293F cells
2days, 8%CO2
cells
lysis buffer
+ sonication
cell lysate
cfg
supernatant
(S)
precipitate
(P)
B
PBS
CLB
RIPA
M-PER
S P S P S P S P
CT52
CT55
CT40
GFP

Figure2

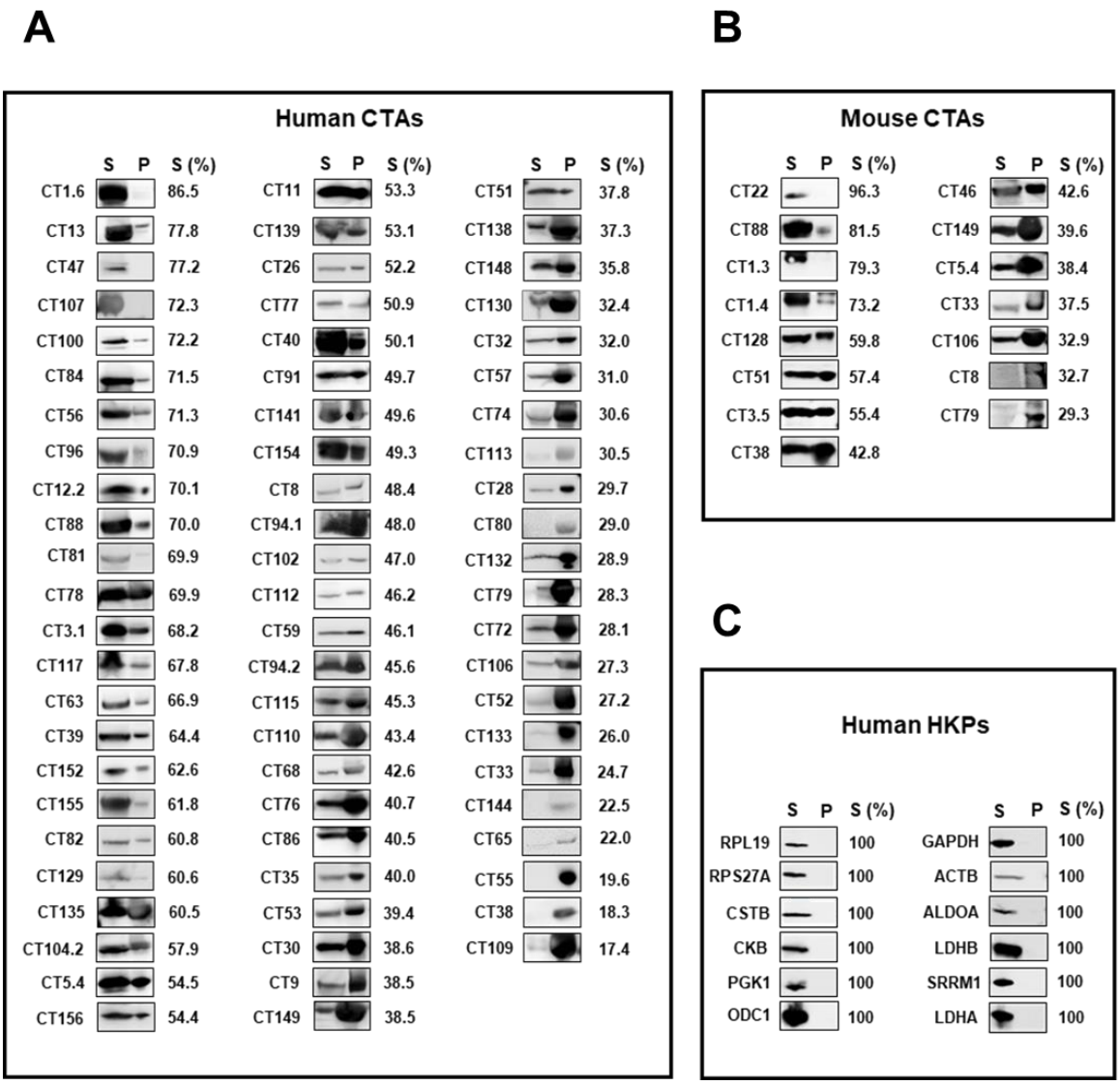

A
Human CTAs
S P S (%)
CT1.6 86.5
CT13 77.8
CT47 77.2
CT107 72.3
CT100 72.2
CT84 71.5
CT56 71.3
CT96 70.9
CT12.2 70.1
CT88 70.0
CT81 69.9
CT78 69.9
CT3.1 68.2
CT117 67.8
CT63 66.9
CT39 64.4
CT152 62.6
CT155 61.8
CT82 60.8
CT129 60.6
CT135 60.5
CT104.2 57.9
CT5.4 54.5
CT156 54.4
CT11 53.3
CT139 53.1
CT26 52.2
CT77 50.9
CT40 50.1
CT91 49.7
CT141 49.6
CT154 49.3
CT8 48.4
CT94.1 48.0
CT102 47.0
CT112 46.2
CT59 46.1
CT94.2 45.6
CT115 45.3
CT110 43.4
CT68 42.6
CT76 40.7
CT86 40.5
CT35 40.0
CT53 39.4
CT30 38.6
CT9 38.5
CT149 38.5
CT51 37.8
CT138 37.3
CT148 35.8
CT130 32.4
CT32 32.0
CT57 31.0
CT74 30.6
CT113 30.5
CT28 29.7
CT80 29.0
CT132 28.9
CT79 28.3
CT72 28.1
CT106 27.3
CT52 27.2
CT133 26.0
CT33 24.7
CT144 22.5
CT65 22.0
CT55 19.6
CT38 18.3
CT109 17.4
B
Mouse CTAs
S P S (%)
CT22 96.3
CT88 81.5
CT1.3 79.3
CT1.4 73.2
CT128 59.8
CT51 57.4
CT3.5 55.4
CT38 42.8
CT46 42.6
CT149 39.6
CT5.4 38.4
CT33 37.5
CT106 32.9
CT8 32.7
CT79 29.3
C
Human HKPs
S P S (%)
RPL19 100
RPS27A 100
CSTB 100
CKB 100
PGK1 100
ODC1 100
GAPDH 100
ACTB 100
ALDOA 100
LDHB 100
SRRM1 100
LDHA 100

Figure3

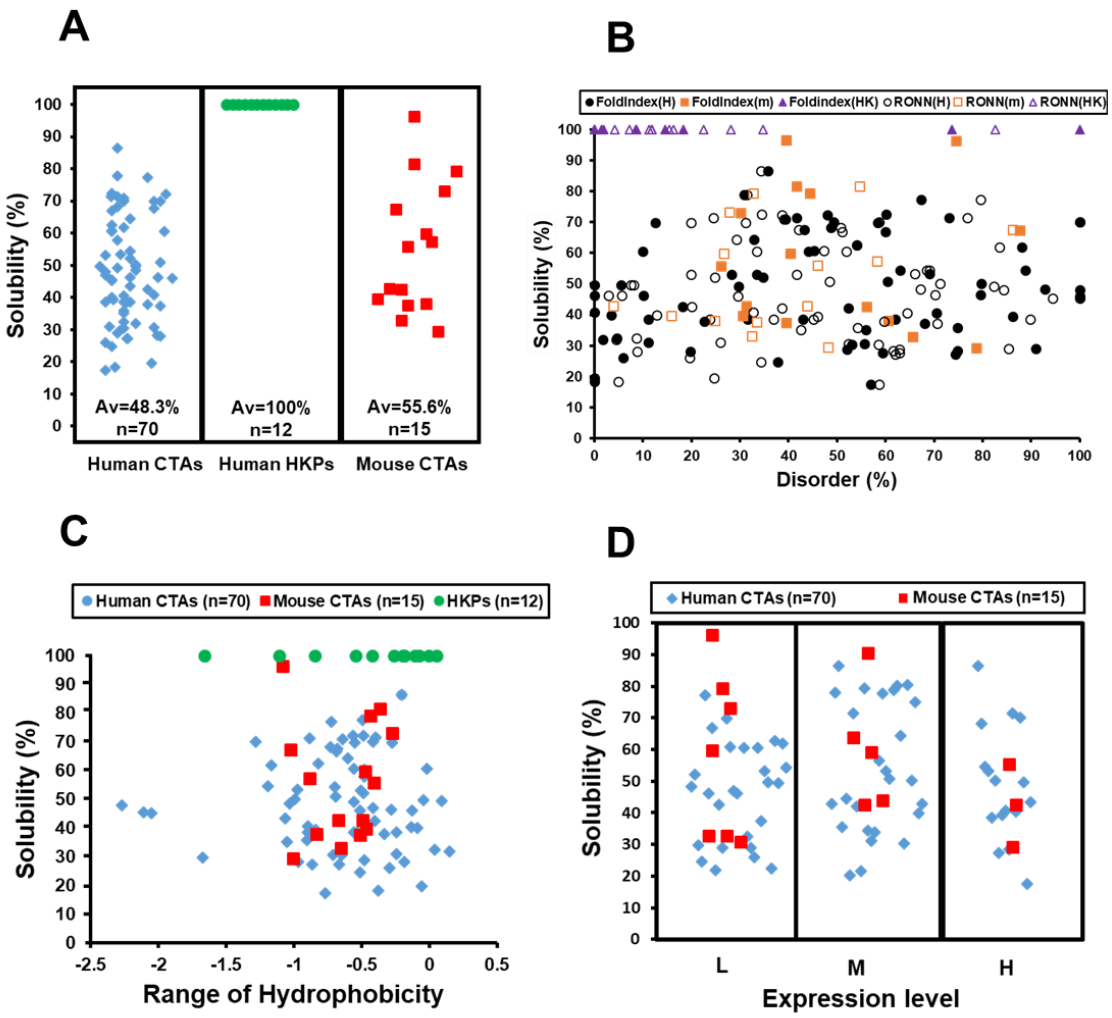
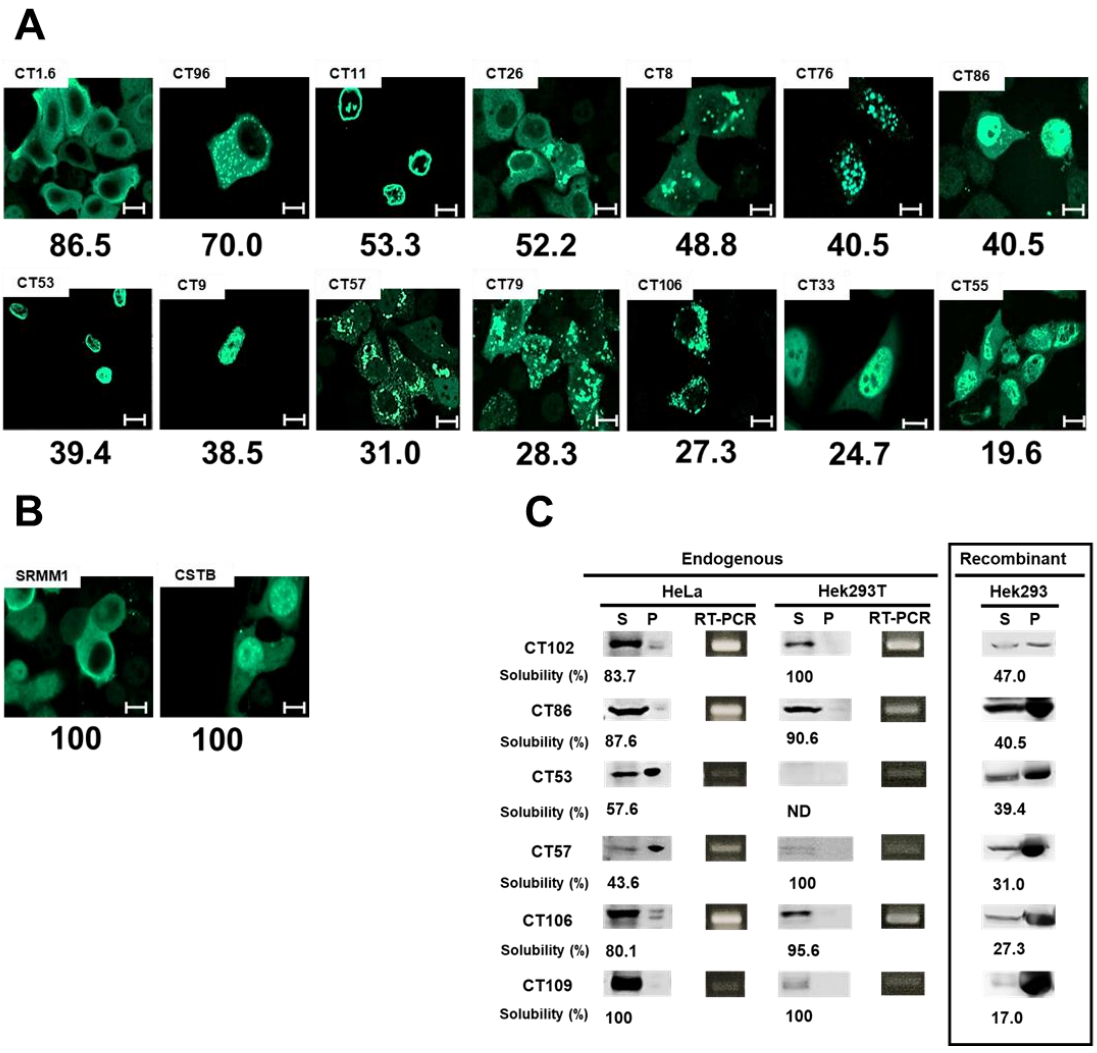

A
Solubility (%)
Av=48.3%
n=70
Av=100%
n=12
Av=55.6%
n=15
Human CTAs
Human HKPs
Mouse CTAs
B
FoldIndex(H) FoldIndex(m) Foldindex(HK) RONN(H) RONN(m) RONN(HK)
Solubility (%)
Disorder (%)
C
Human CTAs (n=70) Mouse CTAs (n=15) HKPs (n=12)
Solubility (%)
Range of Hydrophobicity
D
Human CTAs (n=70) Mouse CTAs (n=15)
Solubility (%)
L
M
H
Expression level

Figure4


A
CT1.6
CT96
CT11
CT26
CT8
CT76
CT86
86.5
70.0
53.3
52.2
48.8
40.5
40.5
CT53
CT9
CT57
CT79
CT106
CT33
CT55
39.4
38.5
31.0
28.3
27.3
24.7
19.6
B
SRMM1
CSTB
100
100
C
Endogenous
Recombinant
HeLa
Hek293T
Hek293
S P
RT-PCR
S P
RT-PCR
S P
CT102
Solubility (%) 83.7
100
47.0
CT86
Solubility (%) 87.6
90.6
40.5
CT53
Solubility (%) 57.6
ND
39.4
CT57
Solubility (%) 43.6
100
31.0
CT106
Solubility (%) 80.1
95.6
27.3
CT109
Solubility (%) 100
100
17.0