Am J Cancer Res 2023;13(5):1863-1883
www.ajcr.us /ISSN:2156-6976/ajcr0150000

Original Article

Structural, topological, and functional characterization of transmembrane proteins TMEM213, 207, 116, 72 and 30B provides a potential link to ccRCC etiology

Joanna Wesoly[1], Natalia Pstrąg[1], Kamil Derylo[3], Barbara Michalec-Wawiórka[3], Natalia Derebecka[1], Hanna Nowicka[1], Arkadiusz Kajdasz[2], Katarzyna Kluzek[2], Malgorzata Srebniak[4], Marek Tchórzewski[3], Zbigniew Kwias[5], Hans Bluyssen[2]

[1]Laboratory of High Throughput Technologies, Adam Mickiewicz University, Poznan, Poland; [2]Laboratory of Human Molecular Genetics, Adam Mickiewicz University, Poznan, Poland; [3]Department of Molecular Biology, Maria Curie-Sklodowska University, Lublin, Poland; [4]Department of Clinical Genetics, Erasmus Medical Center, The Netherlands; [5]Department of Urology and Urological Oncology, Poznan University of Medical Sciences, Poznan, Poland

Received March 6, 2023; Accepted April 2, 2023; Epub May 15, 2023; Published May 30, 2023

Abstract: Due to their involvement in the development of various cancers Transmembrane Proteins (TMEMs) are the focus of many recent studies. Previously we reported TMEM de-regulation in clear cell Renal Cell Carcinoma (ccRCC) with TMEM213, 207, 116, 72 and 30B being among the most downregulated on mRNA level. TMEM down-regulation was also more pronounced in advanced ccRCC tumors and was potentially linked to clinical parameters such as: metastasis (TMEM72 and 116), Fuhrman grade (TMEM30B) and overall survival (TMEM30B). To further investigate these findings, first, we set off to prove experimentally that selected TMEMs are indeed membrane-bound as predicted in silico, we verified the presence of signaling peptides on their N-termini, orientation of TMEMs within the membrane and validated their predicted cellular localization. To investigate the potential role of selected TMEMs in cellular processes overexpression studies in HEK293 and HK-2 cell lines were carried out. Additionally, we tested TMEM isoform expression in ccRCC tumors, identified mutations in TMEM genes and examined chromosomal aberrations in their loci. We confirmed the membrane-bound status of all selected TMEMs, assigned TMEM213, and 207 to early endosomes, TMEM72 to early endosomes and plasma membrane, TMEM116 and 30B to the endoplasmic reticulum. The N-terminus of TMEM213 was found to be exposed to the cytoplasm, the C-terminus of TMEM207, 116 and 72 were directed toward the cytoplasm, and both termini of TMEM30B faced the cytoplasm. Interestingly, TMEM mutations and chromosomal aberrations were infrequent in ccRCC tumors, yet we identified potentially damaging mutations in TMEM213 and TMEM30B and found deletions in the TMEM30B locus in nearly 30% of the tumors. Overexpression studies suggested selected TMEMs may take part in carcinogenesis processes such as cell adhesion, regulation of epithelial cell proliferation, and regulation of adaptive immune response, which could indicate a link to the development and progression of ccRCC.

Keywords: ccRCC, transmembrane protein, subcellular localization, gene expression

Introduction

Renal cell carcinoma (RCC) is a relatively common malignancy accounting for 2% of all adult cancers and causing approximately 100,000 deaths per year wordwide. 80% of RCC cases are classified as clear cell renal cell carcinoma (ccRCC), originating from renal proximal convoluted tubule, with 168,000 new cases annually [1]. Recent incidence of RCC has been steadily increasing, likely due to the higher exposition to risk factors (cigarette smoking, obesity, hypertension) and accidental diagnosis due to improved visualization techniques [2].

On molecular level ccRCC is characterized by de-regulation of genes involved in glycolysis, fatty acid synthesis, mutations in VHL and a number of genes involved in epigenetic regulation (e.g., *PBRM1*, *BAP1* and *KDMC5*) [3-5]. In our previously published microarray-based meta-analysis, we identified a number of ccRCC de-regulated genes significantly represented by transmembrane proteins - TMEMs [6]. TMEMs

are poorly characterized groups predicted *in silico* membrane-integral proteins comprising of more than 300 members. TMEM encoding genes are broadly distributed across the genome, indicating many functional or topological classes. These heterogeneous proteins share very limited sequence homology, and have been suggested to play an important role in cancer [7]. Differential regulation of TMEMs could be observed in many cancers, such as lymphomas (TMEM176) [8], colorectal cancer (TMEM25) [9], meningiomas (TMEM30B) [10], paragangliomas and pheochromocytomas (TMEM127) [11]. A number of TMEM proteins has been functionally assigned as trans-membranous anion channels (e.g. ANO1) [12] and molecules responsible for oncosis (TMEM123) [13], protein glycosylation (TMEM165) [14], pathogen intoxication (TMEM181) [15], as well as innate immunity response (TMEM173) [16]. In case of ccRCC, up- or down-regulation of TMEMs is supported by a number of microarray and RNA-seq data but interestingly, the mechanism of de-regulation and the exact function of majority of TMEMs remains unclear, including their contribution to development and progression of ccRCC.

TMEM213 is the least characterized of the selected TMEMs. TMEM213 was found de-regulated in lung adenocarcinoma and head and neck squamous cell carcinoma [17, 18]. TMEM116 was highly expressed in non-small-cell lung cancer tissues and cell lines, while TMEM207 was formerly linked to gastrointestinal carcinogenesis, oral squamous cell carcinoma and polycystic kidney disease [19-21]. TMEM72 has just recently been shown to associate with Cytoplasmic coat protein complex II and likely take part in membrane trafficking [22]. TMEM30B was described as a phospholipid flippase, required for ER exit of P4-ATP-ases, although its structure has not been previously verified experimentally [23, 24]. TMEM30B expression was also reported to associate with worse overall survival in patients with high-grade serous ovarian cancers [25].

Here, we characterized localization, structure, and topology of the five TMEMs and investigated the occurrence of TMEM isoforms mutations, polymorphisms and chromosomal aberrations in ccRCC tumors derived from Polish patients. We also cloned TMEM213, 207, 116, 72 and 30B and over-expressed full length TMEMs in HEK293 and HK-2 cells in order to elucidate their involvement in cellular processes that could be linked to ccRCC etiology.

Materials and methods

Genetic constructs

TMEM constructs were generated with pEGFP-C1, pEGFP-N1 (Promega) and pcDNA3.1 (Addgene). ORFs of TMEMs were amplified from non-ccRCC kidney with PrimeSTAR GXL DNA polymerase (Clontech) and verified by Sanger sequencing. After purification, DNA fragments were phosphorylated (T4 Polynucleotide Kinase, ThermoFisher Scientific) and ligated (T4 DNA Ligase, ThermoFisher Scientific). The constructs were used to generate cMyc-tagged fusion proteins. A different approach was used to generate TMEM213ORF2, since this ORF differs from TMEM213ORF1 with 3nt, the variant was PCR amplified in two parts.

Cell culture

Human HK-2, HEK293 and monkey COS7 cell lines were grown in Dulbecco's Modified Eagle Medium/Nutrient Mixture F-12 (DMEM-F12) (HK-2) or Dulbecco's Modified Eagle Medium (DMEM) (HEK293 and Cos7) medium, with L-Glutamine supplemented with 10% fetal bovine serum and 1% antibiotic/antimycotic (Biowest), at 37°C with 5% CO_2.

Transfection/TMEM overexpression

Cells were transfected with X-tremeGENE HP DNA Transfection Reagent (Roche) at a ratio of 1 µg plasmid per 2 µl X-tremeGENE HP DNA Transfection Reagent. Cells were harvested/observed under confocal microscope 24-48 h after transfection. RNA samples were measured using NanoDrop and Qubit RNA HS Assay Kit (ThermoFisher Scientific). RevertAid RT Reverse Transcription Kit (Thermo Fisher Scientific) was used for RT. cDNA was measured with Agilent Bioanalyzer High Sensitivity DNA chip (Agilent). The sequencing libraries were prepared with NEBNext Ultra RNA Library Kit for Illumina (NewEngland Biolabs), samples were sequenced on NovaSeq (Illumina), SR50 with 10-15 mln reads per sample. Sequencing quality was verified using fastqc [26] and multiQC [27]. Samples were aligned with STAR [28]

to the human genome version GRCh38. Indexing was assigned with samtools [29, 30] and counts obtained with featureCounts [31]. TopGO (version 2.50.0) and org.Hs.eg.db (version 3.16.0) R packages were used for Gene Ontology (GO) enrichment analysis with differentially expressed genes (|log2FC|≥2, adjusted *P*-value < 0.05) as an input. The selected node size in topGo analysis for biological process was 10, parentchild algorithm and Fischer's exact test with < 0.01 as a cutoff were used [32].

Organelle staining and confocal microscopy

Cell organelles were stained with either CellLight BacMam 2.0 ER-RFP/Golgi-RFP/Plasma Membrane-RFP/Early Endosomes-RFP (ThermoFisher Scientific), ER-Tracker™ Red (BODIPY™ TR Glibenclamide, BODIPY TR Ceramide complexed to BSA or MitoTracker™ Red FM and Hoechst 33342 (Invitrogen). Live-cell microscopy was conducted on LSM780 Zeiss coupled to AxioObserver Z.1 with a Plan-Apochromat 63x/1.40 Oil DIC M27 objective and an environmental chamber. Localization analysis of TMEM proteins was performed in 35 × 10 mm glass bottom dishes (Greiner BioOne) 24 h after transfection. Signal peptide confocal analysis was performed within 24 or 48 h after HK-2, HEK293 or Cos7 transfection. Images were processed with Imaris and presented as maximum intensity projection from all z slices. Live-cell imaging was performed on Nikon A1Rsi with objectives Nikon Plan Apo VC 60x/1.4 Oil DIC N2. For each cell several optical slices of the cell volume were acquired. Time-lapse sequences were acquired through 12-16 h with 2-5 min step between scans. FRAP experiment was conducted as described previously [33].

Fluorescence protease protection (FPP) assay

FPP assay was performed as described previously [34], minimum in duplicate. HK-2 or Cos7 cell were grown on 10-well plate (Greiner Bio-One) and were permeabilized for one minute with saponin (50 ul, 20 uM/ml) and lysed with proteinase K (50 ug/ml) for up to 3 minutes. Movies were recorded using Nikon Eclipse Ti microscope with objective Apo TIRF 100x/1.49 Oil, HQ: GFP filter (Chroma) and Nikon Intensilight illumination system (3 min with 2 s exposition). Fluorescence intensities were measured using Nikon NIS Elements AR software and were normalized to background. CAV1 (ENST00000614113.4) was used as control.

Bioinformatic analysis and signal peptides

NCBI RefSeq sequences of native human TMEMs and Ensembl Transcript IDs were used for bioinformatics analysis: TMEM213 GI: 146229352/ENSG00000214128, the two isoforms: TMEM213ORF1-ENST00000442682.6 and TMEM213ORF2-ENST00000397602.7, and protein isoforms: TMEM213_ORF1 (NP_001078898.1), TMEM213_ORF2 (XP_005250224.1). TMEM207: 46409276/ENST00000354905.2, protein NP_997199. TMEM 116: 30-2058299/ENSG00000198270, isoforms TMEM116ORF1-ENST00000552374.6, TMEM116ORF2-ENST00000550831.7, TMEM116ORF3-ENST00000549537.6 and protein isoforms: NP_001180460.1, NP_001281243, NR_122119. TMEM72: 183227675/ENSG00000187783, isoforms: TMEM72ORF1-ENST00000389583.4 and TMEM72ORF2-ENST00000544540.5 with protein sequences NP_001116848 and NP_001332855. TMEM30B: 63003930/ENST00000555868.1, isoforms: ENST00000554497.1 and ENST00000557163.1. Subcellular localization and cytosolic/nuclear discrimination were predicted using WoLF PSORT [35]. The putative signal sequences and cleavage sites were predicted using: PrediSi [36], SignalBlast [37], PolyPhobius [38], TargetP 1.1 Server [39], SignalP 4.1 Server [38] and WoLF PSORT [35]. The consensus of all programs was used to determine the length of putative signal peptide and the position of possible cleavage site. Plasmids containing putative signal peptide sequence were fused with N- terminal GFP, signal sequence was prolonged with additional sequence of 3 to 9 aa.

Western blot

HK-2, HEK293 or COS7 cells were lysed with RIPA buffer supplemented with SIGMAFAST Protease Inhibitor Cocktail (Sigma), 24hours post transfection. Lysates were sonicated at 4°C. Western blot was performed with GFP (1:1000, Santa Cruz) or α-tubulin antibody (Millipore). Anti-rabbit (1:20 000, Sigma) and anti-mouse (1:2000, Millipore) antibodies were conjugated with horseradish peroxidase, detected with Luminata Forte HRP Substrate (Millipore), visualized with the G: Box (Syngene).

For fluorescence protein visualization on the scanner (TMEM213), samples were prepared without heat denaturation. The gel was scanned directly on Amersham Typhoon RGB with Cy2 Fltr 525BP20, Cy3 Fltr 570BP20, Cy5 Fltr 670BP30 (GE Healthcare).

Patient material, RNA isolation and RT-PCR

Samples were obtained from the Department of Urology and Urological Oncology, Poznan University of Medical Sciences. Research was approved by the Bioethical Committee at UMS (No. 1124/12). All patients granted written informed consent. ccRCC specimens, classified by trained pathologist, were suspended in RNALater (Sigma-Aldrich).

RNA isolation and RT were described previously [6]. qPCR analyses was performed using Bio-Rad CFX Connect system with Bio-Rad iTaq™ Universal SYBR Green Supermix (Bio-Rad Laboratories Inc.). Primers were designed using Primer-BLAST (http://www.ncbi.nlm.nih.gov/tools/primerblast) and Oligo Analyzer 3.1 (https://eu.idtdna.com/calc/analyzer). Forward primer was placed in alternative exon or on the span of two exons flanking the alternative exon, reverse primer in the exon following the alternative exon. When possible, the same reverse primer was used to minimize primer annealing bias. Expression was measured in duplicate, in two separate experiments. Relative incidence of a variant (RIV) coefficient was calculated according to protocol by Londono *et al.* [40]. All calculations were performed in R.

Mutations in TMEM genes

DNA from 84 ccRCC tumors and 12 paired peritumor tissues was isolated with GeneMATRIX Tissue DNA Purification Kit (EurX), 50 ng of DNA was used for PCR. Introns and exons were amplified except for TMEM116 (exons only). The average length of amplicons was 6594 bp (3053 - 10941 bp). Amplicons were purified with Agencourt AMPure XP beads (Beckman Coulter), DNA concentration was measured with Qubit 4 (Thermo Fisher Scientific). 21 amplicons per patient were pooled in 7 nmol concentration, 96 libraries were prepared with NexteraXT kit (Illumina)/TruePrep DNA Library Prep Kit V2 (Vazyme Biotech). Samples were sequenced on MiSeq (Illumina, Reagent Kit v3, PE300) with cluster density of 994±46, clusters PF 85,70%±3.17, Q30 = 71.6%. QC of the run was performed with FastQC [41], adapters and low quality reads were removed with Fastp [42] and aligned to GRCh37.67 with BWA-MEM [43]. appreci8 was used for variant identification [44].

SNP arrays

Copy number B-Allelic frequency analysis was performed using HumanOmniExpress12v1.1 SNP array (733K BeadChip, Illumina Inc). Gene MATRIX Universal DNA Purification Kit (EurX) was used for DNA isolation and 200 ng of DNA was used as an input for the arrays. DNA amplification, tagging and hybridization were performed according to the manufacturer's protocol. The array slides were scanned on a HiScan SQ (Illumina). GenomeStudio version 2011.1 and Nexus BioDiscovery Copy Number 7.0-7.5 with SNP-FASST2 Segmentation were used for data analysis. The HapMap control set provided by the manufacturer served as a control. UCSC built Hg19 (Human Mar. 2009 (NCBI37/hg19) Assembly) was used to analyze the data.

Tumor microenvironment

RNA-Seq FPKM counts (tumors vs. controls) were downloaded from the TCGA (https://www.cancer.gov/tcga). Sequencing and alignment of samples was described previously [5]. The analysis included DEG with $-1<\log_2 FC>1$ expression, FDR ≤ 0.01 (8.335 up-regulated genes and 2.637 down-regulated genes). The immune infiltration scores were generated with CIBERSORT [45]. To decrease redundancy in data sets gene signatures provided by CIBERSORT were intersected with TCGA data. Genes found in both datasets were removed. The immune signature was correlated with a differential gene expression list to identify genes potentially implicated in the attraction of each immune cell type/function. Pearson and Pearson partial correlation were calculated using significantly DEG value with FDR (> 0.3 or < -0.3) and immune signatures.

Results

Intracellular dynamics of TMEMs

To establish if TMEM proteins are indeed integral to membrane, we performed FRAP (Fluorescence Recovery After Photobleaching)

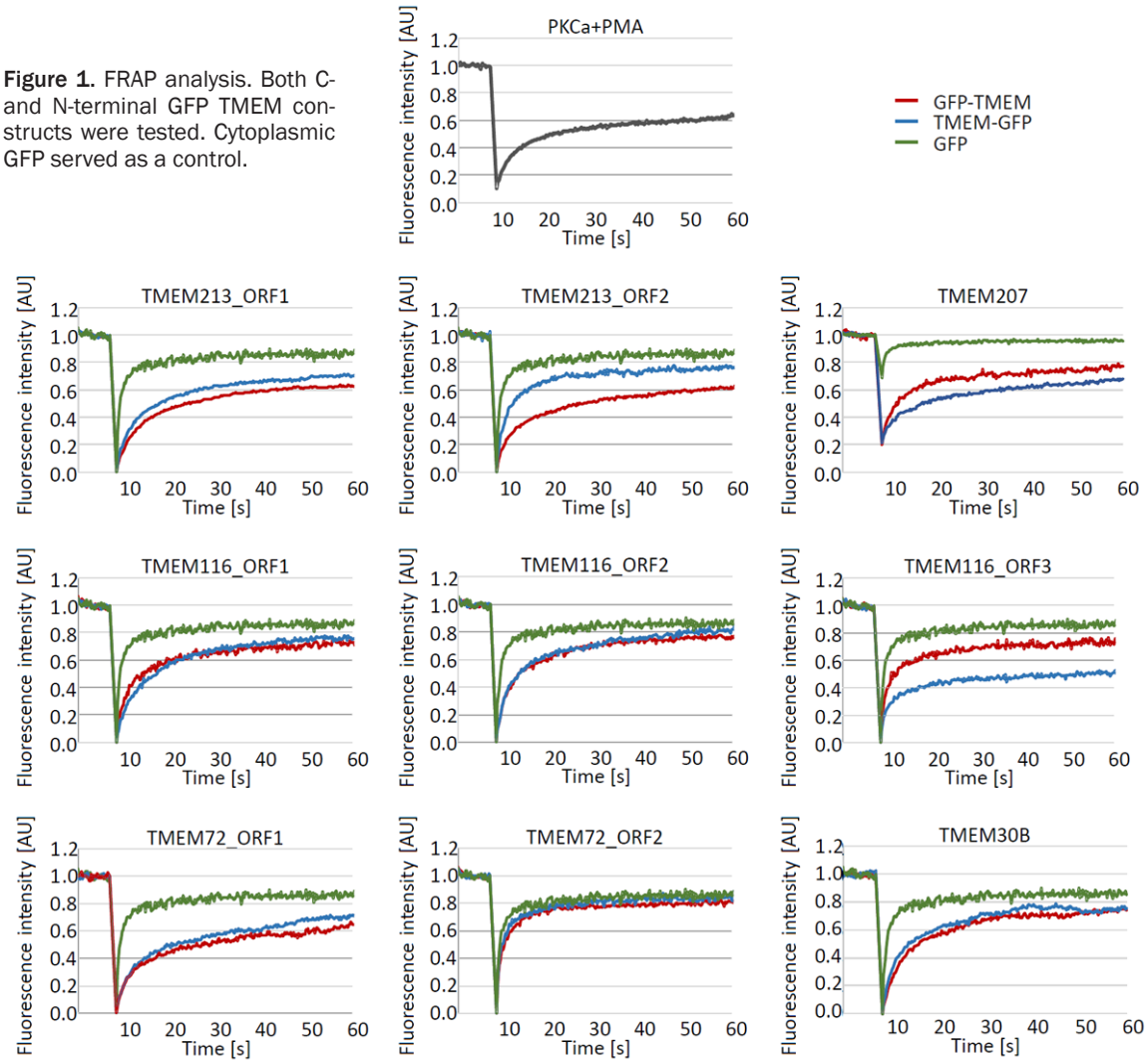
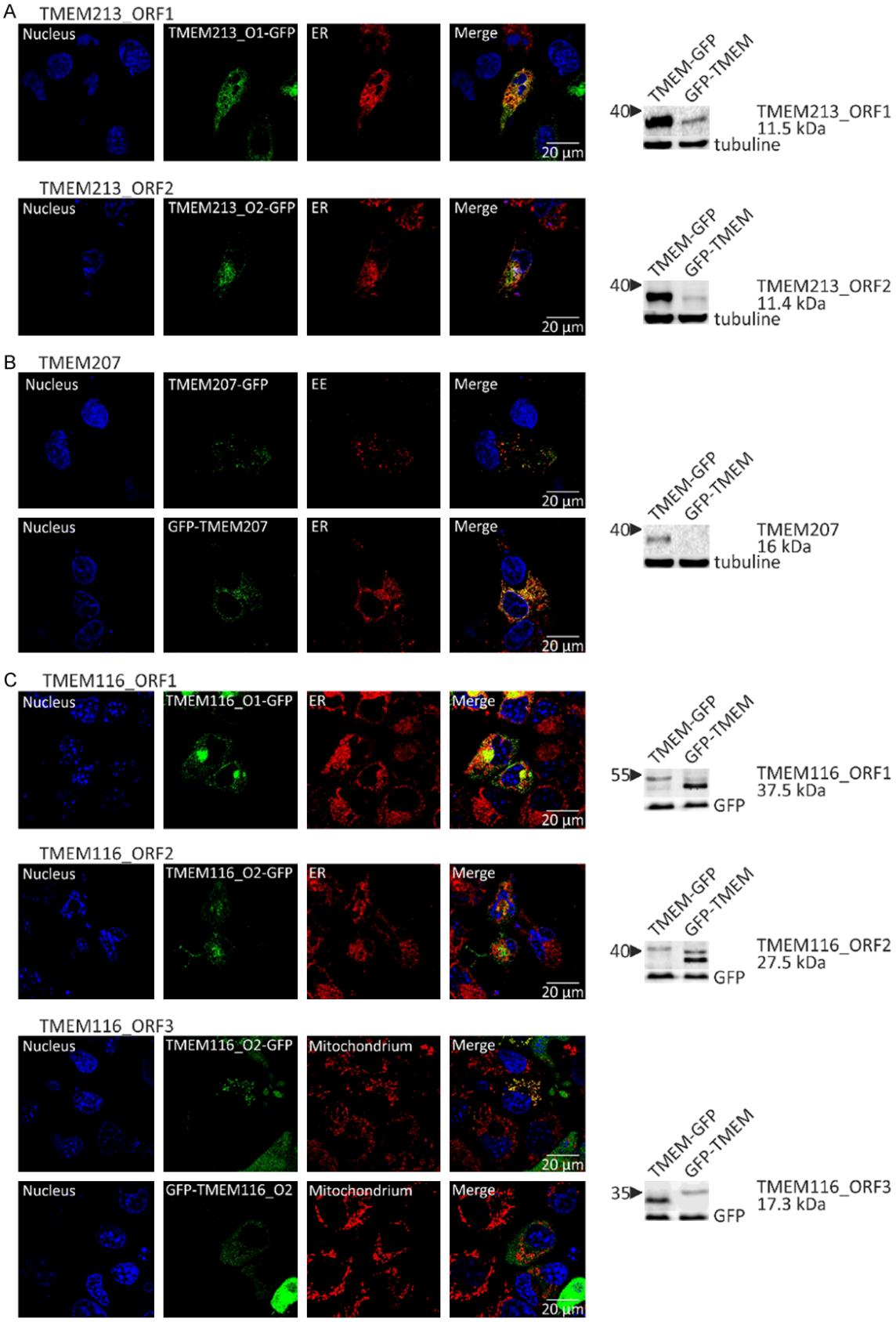


Figure 1. FRAP analysis. Both C- and N-terminal GFP TMEM constructs were tested. Cytoplasmic GFP served as a control.

experiments and analyzed their intracellular dynamics (**Figure 1**). HK-2 or HEK293 cells were transiently transfected with GFP-tagged TMEMs in both, N- and C-terminal, orientations. GFP served as a control of fast-moving protein freely diffusing in the cytoplasm, not bound to any membrane, whereas PKCa activated with PMA was used as a marker of slow-mobility protein immobilized through association with the plasma membrane.

Interestingly, the mobility rate of TMEM213_ORF1 and TMEM213_ORF2 differs significantly. The dynamics of GFP-fused TMEM213_ORF1 was comparable regardless of the orientation of GFP fusion (N- or C-terminal), while TMEM213_ORF2 with N-terminal GFP represented the lowest mobility among the proteins.

The difference in the localization of N- and C-terminal GFP fusions of TMEM207 is also reflected in the dynamics of those proteins. TMEM207 with C-terminal GFP has lower mobility than the same protein with N-terminal GFP fusion. TMEM116_ORF1, TMEM116_ORF2 and TMEM116_ORF3 proteins, with GFP at the N-terminus, are moderately mobile, with the diffusion rate lower than GFP alone. TMEM116_ORF3 with GFP fused at the C-terminus displays mobility that is significantly lower in comparison to other ORFs and similar to mobility rate of PMA-induced PKCa, suggestive of membrane association of this isoform. TMEM72_ORF1, regardless of the GFP fusion, is characterized by low mobility, which is in concordance with our observations as TMEM72_ORF1 is likely bound to EE and to plasma membrane. The high mobility of TMEM72_ORF2 confirms that this isoform is not associated with any membranes. TMEM30B is also characterized by low mobility irrespective of orientation of GFP


A
TMEM213_ORF1
Nucleus
TMEM213_O1-GFP
ER
Merge
20 μm
TMEM-GFP
GFP-TMEM
40
TMEM213_ORF1
11.5 kDa
tubuline
TMEM213_ORF2
Nucleus
TMEM213_O2-GFP
ER
Merge
20 μm
TMEM-GFP
GFP-TMEM
40
TMEM213_ORF2
11.4 kDa
tubuline
B
TMEM207
Nucleus
TMEM207-GFP
EE
Merge
20 μm
Nucleus
GFP-TMEM207
ER
Merge
20 μm
TMEM-GFP
GFP-TMEM
40
TMEM207
16 kDa
tubuline
C
TMEM116_ORF1
Nucleus
TMEM116_O1-GFP
ER
Merge
20 μm
TMEM-GFP
GFP-TMEM
55
TMEM116_ORF1
37.5 kDa
GFP
TMEM116_ORF2
Nucleus
TMEM116_O2-GFP
ER
Merge
20 μm
TMEM-GFP
GFP-TMEM
40
TMEM116_ORF2
27.5 kDa
GFP
TMEM116_ORF3
Nucleus
TMEM116_O2-GFP
Mitochondrium
Merge
20 μm
Nucleus
GFP-TMEM116_O2
Mitochondrium
Merge
20 μm
TMEM-GFP
GFP-TMEM
35
TMEM116_ORF3
17.3 kDa
GFP

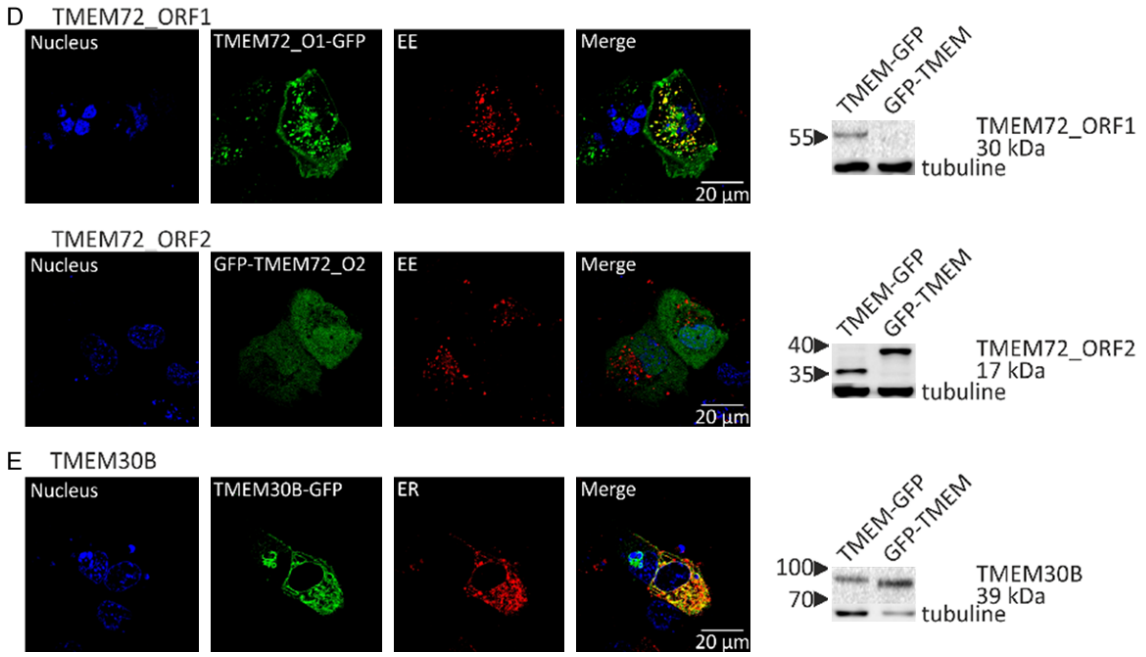


Figure 2. Cellular localization of GFP-tagged TMEMs. Photos depict the results with TMEM-GFP construct, Western blot analysis included.

fusion, which suggests that TMEM30B is a membrane-bound protein. Our results confirm membrane binding of all analyzed TMEMs, except from TMEM72_ORF2 present in cytoplasm and nucleoplasm.

Subcellular localization of TMEMs

To analyze the subcellular localization of TMEM proteins their GFP-fused variants were obtained. All TMEMs were prepared in two orientations, as N- and C-GFP fused proteins. GFP-TMEM/TMEM-GFP expression was verified using standard or fluorescent Western blot (**Figure 2**). TMEM213, 207, 72 and 30B fusion proteins were of expected size, although with N-terminal GFP fusion we did not detect a band or only a weak band for TMEM213_ORF1, TMEM213_ORF2, TMEM207, and TMEM72_ORF1. This is in line with the probable presence of a cleavable signal peptide. Additionally, we did not detect the signal corresponding to C-terminal GFP of TMEM72_ORF2, which introduces a possibility of processing of TMEM72_ORF2 C-terminus. Expression of three isoforms of TMEM116 was very low and detectable only after direct SDS gel scan (**Figure 2C**). There is always a possibility that GFP tag (26,9 kDa) could interfere with the expression or cellular localization of analyzed proteins, especially those of relatively small size like TMEMs (11,5-39 kDa), therefore we additionally cloned all isoforms into pcDNA-3.1/cMyc, and overexpressed their cmyc-fused forms in HK-2 or COS7 cells and confirmed our results (data not shown).

TMEM localization was tested in nucleus, endoplasmic reticulum (ER), early endosomes (EE), mitochondria, Golgi apparatus and plasma membrane. Both ORF variants of TMEM213 localize in ER (**Figure 2A**). Full length TMEM207 fused with GFP at its C-terminus was identified in early endosomes (**Figure 2B**), however when GFP was present on its the N-terminus was localized in ER, which may be explained by the presence of non-processed protein, still in ER, before the signal peptide cleavage. Two longer isoforms of TMEM116 (ORF1, ORF2) were detected in ER, whereas the shortest TMEM116 isoform (ORF3) was found in mitochondria (**Figure 2C**), likely due to the presence of putative MLS on its N-terminus. We were not able to determine the exact localization of TMEM116_ORF3 in the mitochondrion (outer vs. inner membrane). Interestingly, GFP fusion on C-terminus of this shortest isoform interferes with the end-point protein localization, since GFP-TMEM116_ORF3 localizes in the cytoplasm. Next analyzed protein was TMEM72

with its two isoforms. TMEM72_ORF1 was detected in early endosomes with a fraction attributed to the plasma membrane, while its shorter variant, TMEM72_ORF2, was distributed throughout the cytoplasm and the nucleus (**Figure 2D**), likely due to the loss of transmembrane domains and presence of predicted nuclear localization signal (NLS) within its amino-acid sequence. TMEM30B was clearly confined to ER (**Figure 2E**), in agreement with previous results obtained in THP-1 and U2OS cells [23, 46].

We also performed immunohistochemistry with c-myc tagged TMEMs and confirmed the presence of TMEM213 (both isoforms), TMEM116 (ORF_1, ORF_2) and TMEM30B in ER, TMEM207 localized in EE and TMEM72 in EE and plasma membrane (data not shown).

Putative signal peptide cloning, overexpression, and cellular localization

Since many secretory or membrane proteins typically possess 15-30 aa long N-terminal signal peptide (SP), we investigated the presence and location of such sequences. SP directs newly synthesized proteins to ER and to the target organelle. WoLF PSORT predicted the presence of cleavage sites for all TMEMs, while SP was predicted only for TMEM116_ORF2 (1-42aa). Interestingly, TargetP 1.1 Server predicted SP additionally for TMEM116_ORF1 (1-16aa) and TMEM72 (1-22aa) and a mitochondrial targeting peptide for TMEM116_ORF3 (**Figure 3A-D**).

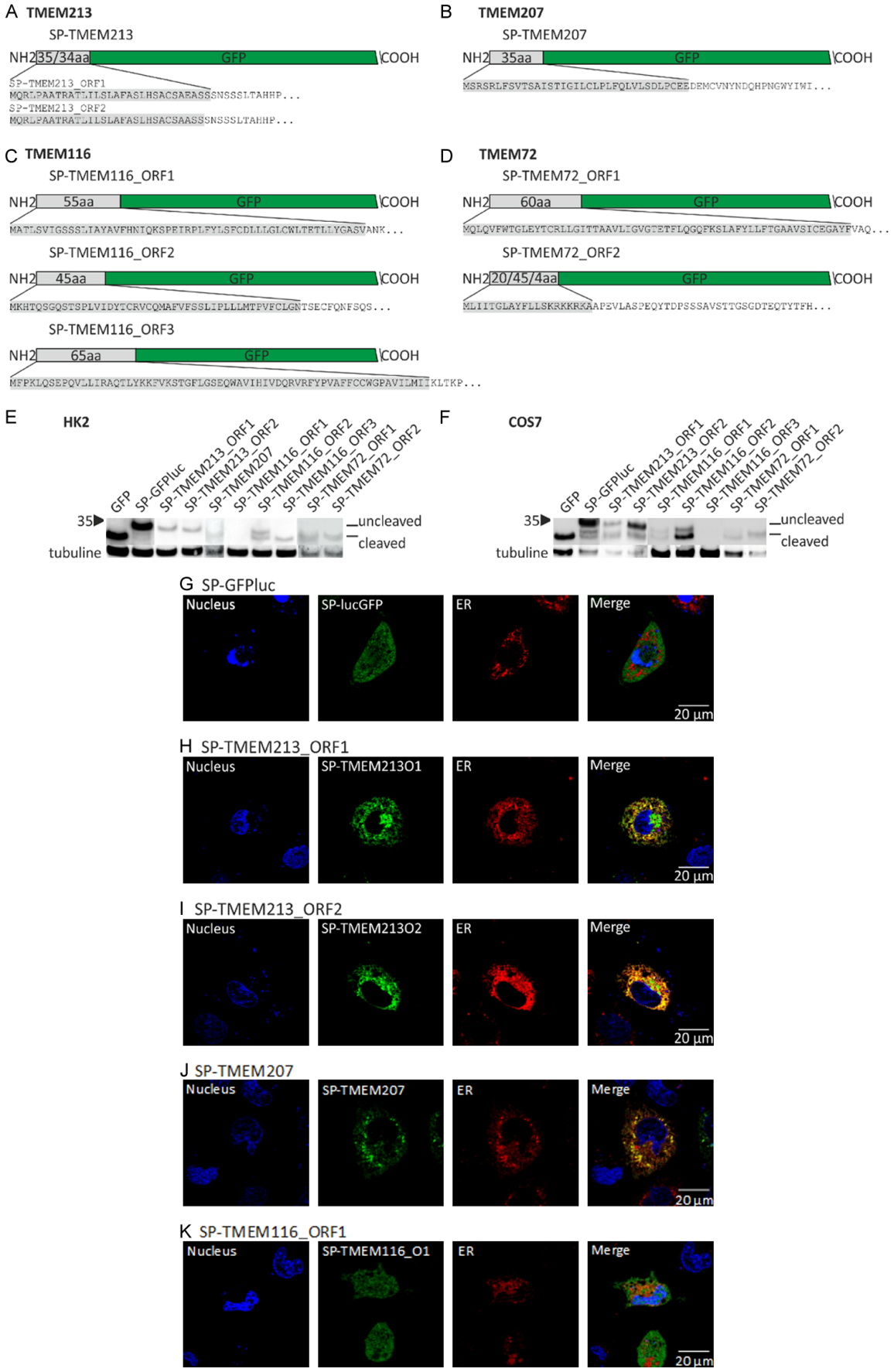
Localization of putative cleavage sites was also predicted by PrediSi, SignalBlast, and SignalP 4.1 Server, however their positions differed. The most convergent results were obtained for TMEM213_ORF1 (position 27aa), TMEM213_ORF2 (position 27 or 28aa), TMEM207 (position 29aa), and TMEM30B (52aa, SP not predicted).

To experimentally validate the presence and expression of SPs, putative sequences were cloned, fused with GFP, and overexpressed in HK-2 and COS7 cells (**Figure 3E**, **3F**). Interestingly, SP_TMEM213_ORF1 and SP_TMEM213_ORF2, appear not to be processed and the bands correspond to proteins longer than GFP alone in HK-2 cells. In COS7 cells we observe two bands, likely corresponding to processed and unprocessed proteins, which may indicate that recognition and processing of SP and use of cleavage sites may be cell type dependent. The results of Western blot (**Figure 3E**, **3F**) indicate that proteins encoded by SP_TMEM207, SP_TMEM116_ORF3, SP_TMEM72_ORF1 and SP_TMEM72_ORF2 have the same length as GFP, suggestive of protein processing and cleavage of SP. Two bands were detected for SP_TMEM116_ORF2 in HK-2 cells, while SP_TMEM116_ORF1 is detectable only in COS7 cells, again with two bands. SP_TMEM72_ORF2 appears not to be processed in COS7 cells or to contain a SP.

Overexpression of obtained constructs revealed that SP-TMEM72_ORF1, SP-TMEM116_ORF2, SP-TMEM207, SP-TMEM213_ORF1 and SP-TMEM213_ORF2 localize in the ER, suggesting that used sequences can drive GFP localization (**Figure 3G-J**, **3L**). However, full-length TMEM72_ORF1 and TMEM207 localize in EE, so we assume that there must be additional cis or trans factor that directs these proteins from ER to EE. Exclusively cytoplasmic localization of SP-TMEM72_ORF2 is surprising because the full-length protein has cytoplasmic and nuclear distribution. Additionally, overexpression of SP-TMEM72_ORF2 short or SP-TMEM72_ORF2 long did not result in the enrichment of nuclear distribution as expected, and we observed a distribution typical for GFP which indicates that used sequences were not sufficient. Our results may imply that TMEM72_ORF2, despite possessing NLS, does not carry a typical SP, and only full-length protein can drive its localization (**Figure 3N**, **3O**). The lack of ER and mitochondrial localization of SP-TMEM116_ORF1 and SP-TMEM116_ORF3 respectively, may demonstrate that TMEM sequences used were not sufficient to change the localization of GFP (See **Figure 3K**, **3M**).

TMEM orientation in the cellular membrane

Fluorescence protease protection (FPP) assay was utilized to determine the orientation of TMEM proteins in the membranes. Membrane embedded Caveolin 1 (Cav1) with cytoplasmic N- and C-terminus and cytoplasmic GFP served as controls. The decay of GFP signal after incubation with proteinase K was monitored for up to 3 minutes. N-terminus of TMEM213 appears to be exposed to the cytoplasm with maximal


A TMEM213
SP-TMEM213
NH2
35/34aa
GFP
COOH
SP-TMEM213_ORF1
MQRLPAATRATLILSLAFASLHSACSARASSSNSSSLTAHHP...
SP-TMEM213_ORF2
MQRLPAATRATLILSLAFASLHSACSAASSSNSSSLTAHHP...
B TMEM207
SP-TMEM207
NH2
35aa
GFP
COOH
MSRSRLFSVTSAISTIGILCLPLFQLVLSDLPCEEDEMCVNYNDQHPNGWYIWI...
C TMEM116
SP-TMEM116_ORF1
NH2
55aa
GFP
COOH
MATLSVIGSSSLIAYAVFHNIQKSPEIRPLFYLSFCDLLLGLCWLTFTLLYGASVANK...
SP-TMEM116_ORF2
NH2
45aa
GFP
COOH
MKHTQSGQSTSPLVIDYTCRVCQMAFVFSSLIPLLLMTPVFCLGNTSECFQNFSQS...
SP-TMEM116_ORF3
NH2
65aa
GFP
COOH
MFPKLQSEPQVLLIRAQTLYKKFVKSTGFLGSEQWAVIHIVDQRVRFYPVAFFCCWGPAVILMIIKLTKP...
D TMEM72
SP-TMEM72_ORF1
NH2
60aa
GFP
COOH
MQLQVFWTGLEYTCRLLGITTAAVLIGVGTETFLQGQFKSLAFYLLFTGAAVSICEGAYFVAQ...
SP-TMEM72_ORF2
NH2
20/45/4aa
GFP
COOH
MLIITGLAYFLLSKRKKRKAAPEVLASPEQYTDPSSSAVSTTGSGDTEQTYTFH...
E HK2
GFP
SP-GFPluc
SP-TMEM213_ORF1
SP-TMEM213_ORF2
SP-TMEM207
SP-TMEM116_ORF1
SP-TMEM116_ORF2
SP-TMEM116_ORF3
SP-TMEM72_ORF1
SP-TMEM72_ORF2
35
tubuline
uncleaved
cleaved
F COS7
GFP
SP-GFPluc
SP-TMEM213_ORF1
SP-TMEM213_ORF2
SP-TMEM116_ORF1
SP-TMEM116_ORF2
SP-TMEM116_ORF3
SP-TMEM72_ORF1
SP-TMEM72_ORF2
35
tubuline
uncleaved
cleaved
G SP-GFPluc
Nucleus
SP-lucGFP
ER
Merge
20 μm
H SP-TMEM213_ORF1
Nucleus
SP-TMEM213O1
ER
Merge
20 μm
I SP-TMEM213_ORF2
Nucleus
SP-TMEM213O2
ER
Merge
20 μm
J SP-TMEM207
Nucleus
SP-TMEM207
ER
Merge
20 μm
K SP-TMEM116_ORF1
Nucleus
SP-TMEM116_O1
ER
Merge
20 μm

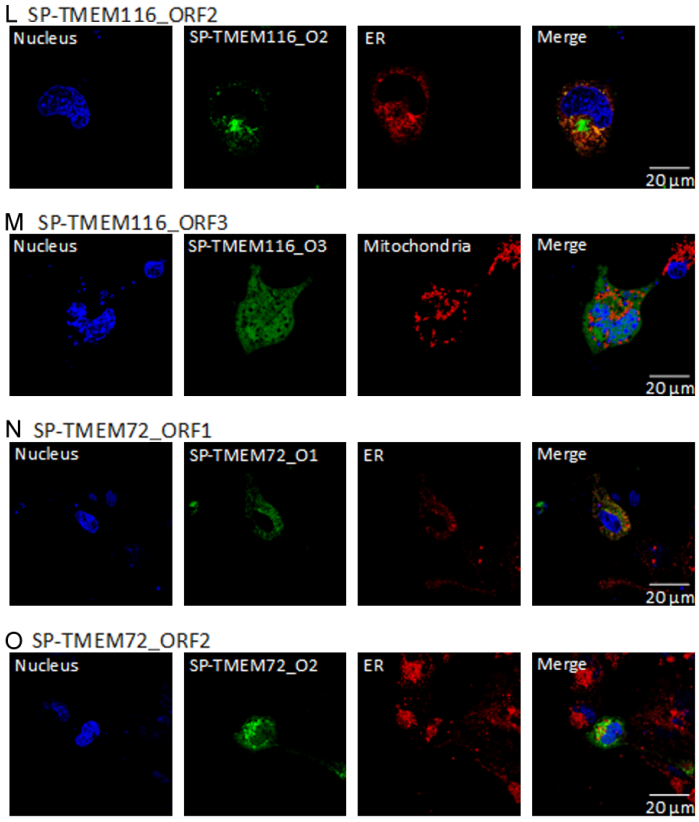


Figure 3. Analysis of TMEM signaling peptides (SP) and mitochondrial localization signal (MLS). (A-D) Putative sequence of TMEM signaling peptides and MLS. Expression of SP/MMLS GFP constructs in HK-2 (E) and COS7 cells (F). The presence of a shorter band on the Western blot likely represents the protein with cleaved SP. (G-O) Cellular localization of TMEM-specific SP/MLS-GFP.

fluorescence (F_{max}) at 135sec, while C-terminus faces ER lumen (F_{max}≥240sec). TMEM207 holds cytoplasmic C-terminus (F_{max}=165sec), and N-terminus is directed towards EE lumen (F_{max}=190sec). C-termini of ORF1 and ORF2 of TMEM116 face the cytoplasm (F_{max}=105sec, respectively), while N-termini are within ER lumen (F_{max}=110sec). The shortest TMEM-116 isoform has internalized N-terminus (F_{max}=100sec), while C-terminus appears to face the cytoplasm (F_{max}=80sec), which is not in line with the presence of two transmembrane domains as predicted, and this finding requires further examination. C-terminus of TMEM72 faces the cytoplasm (F_{max}=140sec), while the N-terminus is protected in EE lumen (F_{max}=240sec). Similar fluorescence decay of both ends of TMEM30B suggests both are cytoplasmic (F_{max}=180sec). In most cases we observed discrepancies between predicted and experimental models as depicted in **Figure 4**.

TMEM overexpression and gene ontology analysis

Since selected TMEMs are expressed on a low level in cell lines of epithelial origin and knock-down experiments are not feasible, we decided to overexpress them in HEK293 and HK-2 cells in order to learn in which cellular processes they could be involved in. Overexpression studies were performed in duplicate, and expression of selected genes was validated in HEK293 and HK-2 cell lines using Real-Time PCR (data not shown). The complete list of TMEM de-regulated genes is available in Supplementary Table 1A. Gene ontology analysis was performed with topGO (**Figure 5**). The majority of the identified pathways were shared among the genes de-regulated by TMEM overexpression and included regulation of cytosolic ion concentration, cellular developmental processes, extracellular structure organization, nucleosome organization and assembly etc.

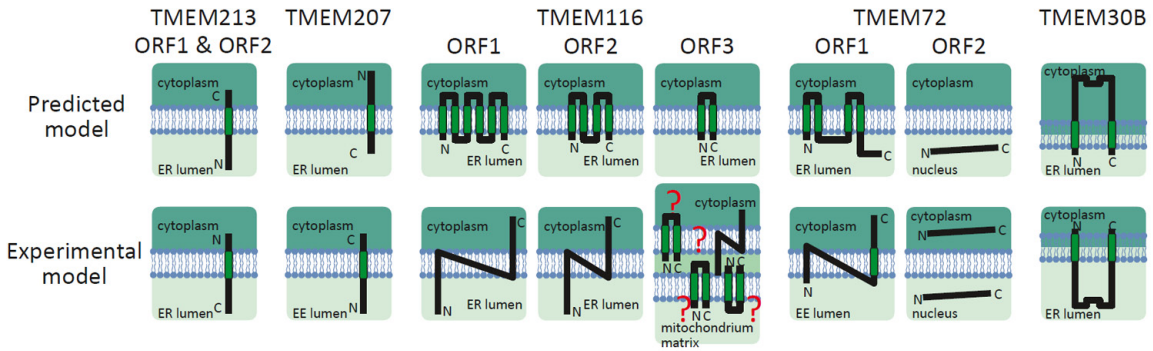


Figure 4. Schematic comparison of predicted vs. experimental data concerning TMEM orientation. The figure represents an updated analysis from Wrzesinski *et al.* with WoLF PSORT and new online tools: PrediSi, SignalBlast, PolyPhobius, TargetP 1.1. Server and SignalP 4.1. Server.

Interestingly TMEM213 overexpression led to enrichment of de-regulated genes grouped as transmembrane transport that included 24 members of Solute carrier family (SLC), ten voltage-gated potassium channels (KCNA) and ATPases: ATP2B3, ATP6V0D1 and ATP6V1C1. Genes de-regulated by TMEM207 overexpression exclusively, were grouped in the pathway of response to hypoxia: HIF3A, PTEN, MYC, TGFB2, TGFB3 and others. Negative regulation of angiogenesis was assigned specifically to TMEM116 overexpression exemplified by ISM1, THBS4 and TNMD genes.

TMEM72 de-regulated genes were uniquely allocated to protein phosphorylation pathway represented by MAPK13, MAPK15 and MAPK4 (See Supplementary Table 1B).

We were particularly interested in genes de-regulated explicitly by overexpression of individual TMEMs, and found 43 TMEM213-specific genes, 148 TMEM207-specific, 88 TMEM116, 34 TMEM72, and 70 TMEM30B-specific genes. In case of TMEM213, the most interesting candidates include ATP6V1C1 (↓2,04), cilium-localized GPR19 (↓2,31), transcription factors: HOXA7 (↓2,18), ZNF506 (↑2.06), ZNF785 (↓2,01) and NQO1 (↓2,16) de-regulated in many tumors. Most notably, TMEM207-specific genes represented Solute Carrier Family: SLC13A3 (↓2,07), SLC25A24 (↓2,11), SLC25A45 (↓2,38), SLC30A10 (↑2,03), immune modulation: FOXI3 (↑2,08), IKZF2 (↓2,22) and tumorigenesis: IQANK1 (↓2,41), KLK10 (↑2,30), KRCC1 (↓2,90). Interestingly, TMEM116-specific genes included ERG (↑2,06), RUNX3 (↑2,05) and TNK1 (↑2,05) implemented in tumorigenesis, cilium function: CFAP221 (↓2,05), PLA2G3 (↑5,13) and kidney function: CLCNKB (↓2,73). TLR1 (↓5,54) is the most interesting TMEM72-specific gene, although its overexpression led to de-regulation of cell cycle genes as well: FRK (↓2,84), GADD45B (↑2,02). Metalloproteinase MMP17 (↓7,28) accompanied by MMP inhibitor TIMP3 (↑2,01), proto-oncogene PIM3 (↑2,01) and sodium channel SCNN1B (↑3,39) represent TMEM30B-specific genes.

In attempt to elucidate TMEM involvement in the etiology of ccRCC we cross-referenced names of genes de-regulated in ccRCC using TCGA RNAseq data with overexpression data derived from HEK293. As a result of GO analysis genes assigned to five general categories: developmental processes, cellular developmental processes, regulation of developmental processes, cell adhesion and cell projection processes were shared throughout all data sets.

Genes grouped into cellular protein modification process, glycosylation, negative regulation of vasculature development and blood vessel morphogenesis and regulation of epithelial cell proliferation were shared exclusively with TMEM213 (See Supplementary Table 1C).

Most significantly cell junction assembly and regulation of adaptive immune response, together with cellular carbohydrate biosynthetic process were assigned to TMEM207. Genes de-regulated by TMEM116 overexpression common with ccRCC de-regulated genes were collected into cell population proliferation, protein modification process, insulin-like growth factor receptor signaling pathway and negative regulation of hormone metabolic process.

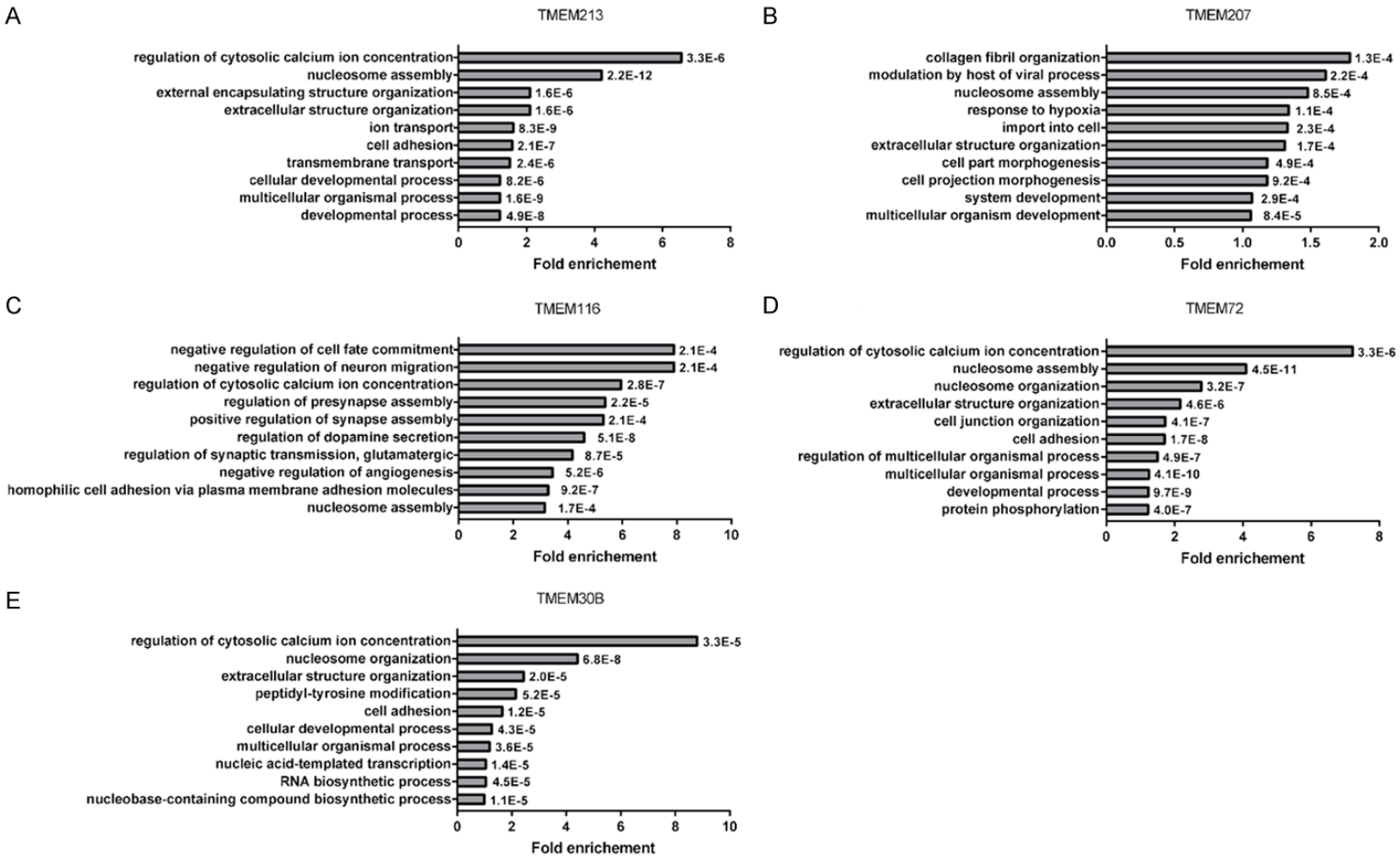


Figure 5. Functional analysis of TMEM de-regulated genes using topGO.

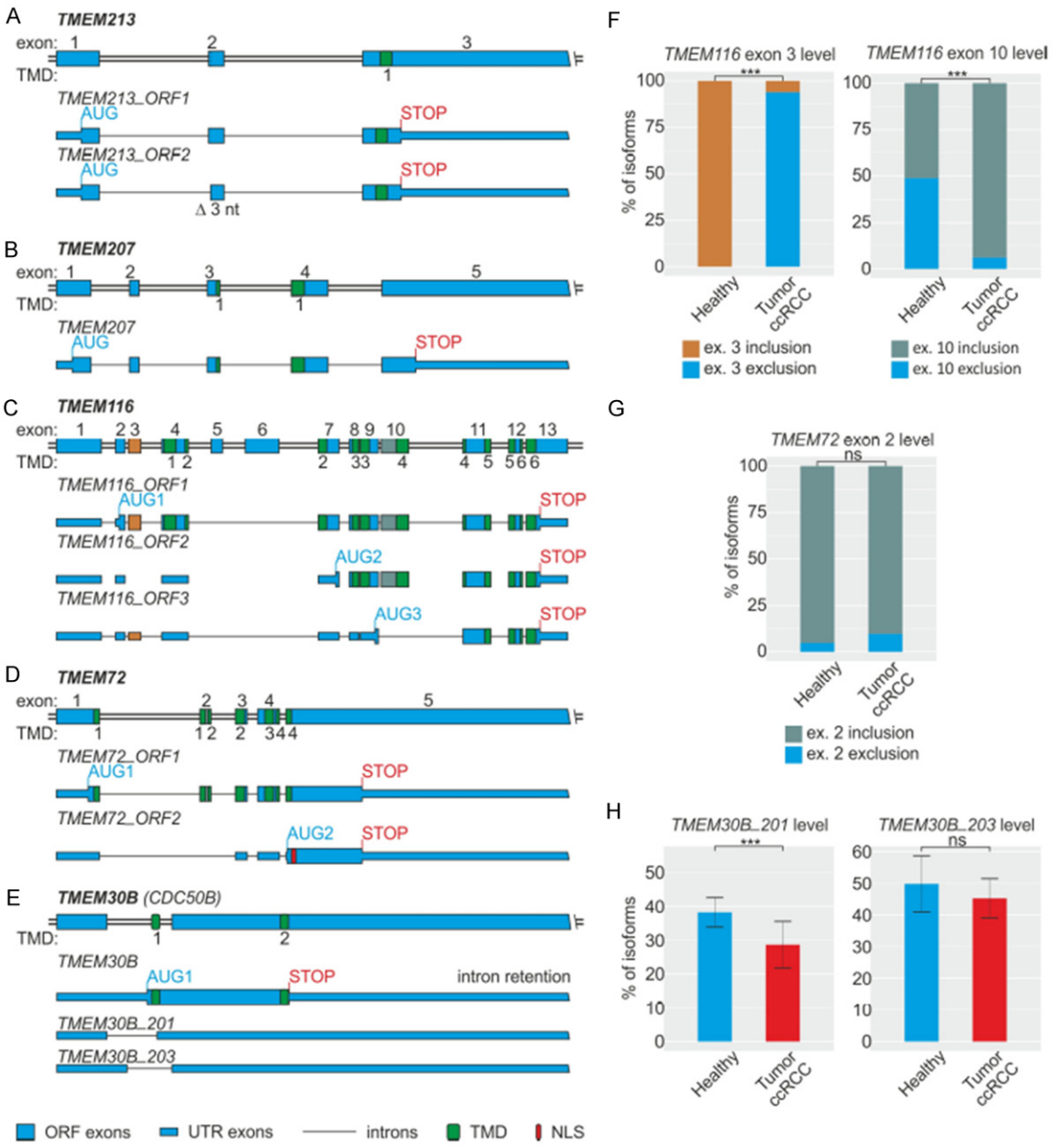


Figure 6. Schematic representation of the structure of investigated TMEM genes (A-E). TMEM isoform expression in ccRCC tumors derived from Polish patients (F-H).

Most notably, protein phosphorylation and glutathione metabolic process characterized genes common with TMEM72 overexpression. TMEM30B overexpression and TCGA data shared genes assigned to regulation of transcription, DNA-templated, regulation of RNA metabolic process, positive regulation of steroid biosynthetic process.

All in all, these observations, although indirect, give us an indication of potential involvement of selected TMEMs in the processes de-regulated in ccRCC.

TMEM mRNA isoforms

Exon composition of *TMEM* mRNAs was determined using ENSEMBL, NCBI, UCSC and ISOexpresso. *TMEM213* has two mRNA isoforms due to alternative use of 3' splice site in intron 1 (**Figure 6A**). In consequence exon 2 contains or lacks the AAG codon. The protein

isoform TMEM213_ORF1 is 107aa long and TMEM213_ORF2 lacks lysine at the position of 27aa. Repeat-rich gene structure of *TMEM213* and three nucleotide difference in the sequence rendered it impossible to detect both isoforms with qPCR. Only one *TMEM207* mRNA isoform encoding a protein of 146aa has been reported (**Figure 6B**), and *TMEM207* is expressed in kidney and down-regulated in ccRCC as shown previously [6]. Three *TMEM116* isoforms were expressed in kidney and ccRCC tumors: *TMEM116_ORF1*, *TMEM116_ORF2* and *TMEM116_ORF3* (**Figure 6C**). The longest *TMEM116* transcript (TMEM116_ORF1) encodes a protein of 337aa, 245aa variant (TMEM116_ORF2) could be translated from a second AUG codon located in exon 7 as a result of exon 3 exclusion. Exclusion of exon 10 led to utilization of the third AUG codon in exon 9 of *TMEM116* and generation of a transcript spanning 151aa (TMEM116_ORF3). In ccRCC *TMEM116* exons 3 and 10 are alternatively spliced, and exon 3 inclusion is common in healthy tissue, though it is excluded in up to 5% of ccRCC tumors (P = 8.054e-05) (**Figure 6F**). Exon 10 inclusion in controls reached 50%, while in ccRCC significantly increased up to 95% (P = 3.553e-05) (**Figure 6F**).

TMEM72 transcript undergoes alternative splicing of exon 2 (**Figure 6D**), and in consequence two isoforms are generated: *TMEM72_ORF1* and *TMEM72_ORF2*, encoding proteins of 275aa and 157aa, respectively. Exclusion of exon 2 leads to the loss of all predicted transmembrane domains (**Figure 6D**, lower panel). In ccRCC tumors, exon 2 exclusion occurs more frequently as compared to controls (6%) but does not reach statistical significance (P > 0.05) (**Figure 6G**).

Differential primary transcript processing has been reported for *TMEM30B* (*CDC50B*) (**Figure 6E**). First variant encodes the full length 351aa long protein. The remaining two isoforms are a consequence of intron removal (*TMEM30B_201*, *TMEM30B_203*). These isoforms share a start codon but have different length of 5' and 3' ends. Both isoforms (*TMEM30B_201, TMEM30B_203*) do not contain an ORF, and are not translated. Frequency of both mRNAs decreases in ccRCC tumors as compared to control tissue (**Figure 6H**), although only in case of *TMEM30B_201* this alteration is statistically significant (P = 7, 07e-05).

Mutations in TMEM genes

To analyze if potential de-regulation of TMEMs occurs due to the presence of mutation in TMEM genes we performed amplicon sequencing which yielded 12, 09 GB of data, with read lengths ranging from 145 to 250 bp (mean: 189 bp). All 96 samples were included in downstream analysis with 0.1-0.4 million aligned reads per sample. We found previously reported benign polymorphisms in *TMEM213*, *207*, *116* and *30B* (See **Table 1**). No sequence variants were detected in *TMEM72*. Interestingly, two potentially damaging nucleotide substitutions were identified in ccRCC tumors in *TMEM213* (rs780125224) and *TMEM30B*. The MutPred tool from Ensembl Variant Effect Predictor classifies rs780125224 in *TMEM213* as "Loss_of_helix" in the transmembrane domain, although with a weak statistical strength (P = 0.3949). A/C substitution in TMEM30B (chr. 14, position 61747114) leads to replacement of valine by glycine localized in the helical part of exoplasmic loop and is assigned as damaging or possibly damaging, likely causing loss of protein stability (P = 0.0151). As a finding it is particularly notable since this mutation has not been reported previously.

Chromosome rearrangements of TMEM loci

In order to investigate if the down- regulation of TMEM expression in ccRCC tumors is caused by presence of chromosomal rearrangements we performed targeted high resolution chromosomal aberration analysis of five investigated TMEMs in 50 tumors and 12 control samples (previously published GWAS data) [46]. Control tissues, collected from surgical margin and identified histopathologically as pathologically unchanged did not contain any genomic copy number variants (CNVs) > 0.5 Mb. As shown in **Table 2** TMEM213 locus was unaltered in 86% of the samples, with only six tumors displaying, interestingly, gains and one - LOH. Similarly chromosomal rearrangements were infrequent in TMEM207, 116 and 72 with 82%, 92% and 92% samples, respectively, with no acquired CNVs. Losses in TMEM30B locus were observed in 14 ccRCC tumors (28%), which

Table 1. List of identified variants in TMEM genes

Gene	Number of samples with SNP	Chr	Pos	Ref	Alt	Type	Mutation	VAF (this study)	EUR_AF	ClinPred_ pred	ClinPred_ rankscore	ClinPred_ score	dbSNP	Cosmic
TMEM213	1	7	138487711	C	T	missense_variant	gCg/gTg|p.Ala74Val	0,0104	-	Damaging	0,53727	0,886820972	rs780125224	COSV100022911/COSM8205530
TMEM213	81	7	138486096	G	C	missense_variant	aGc/aCc|p.Ser36Thr	0,5833	0,6282	Tolerated	0,0001	0,001069661	rs61729750	COSV59473104/COSM3762403
TMEM207	1	3	190167534	G	A	missense_variant	cCg/cTg|p.Pro22Leu	0,0104	0,002	Tolerated	0,00436	0,016818344	rs148503262	-
TMEM207	20	3	190158168	G	C	missense_variant	Ctg/Gtg|p.Leu57Val	0,1041	0,1441	Tolerated	0,00525	0,018002078	rs35161724	-
TMEM116	1	12	112374634	C	G	missense_variant	aaG/aaC|p.Lys150Asn	0,0052	-	Tolerated	0,22335	0,23816872	rs966536180	COSV61392769/COSM357502
TMEM116	25	12	112375990	A	C	missense_variant	Tgt/Ggt|p.Cys-114Gly	0,1354	0,165	Tolerated	0,00006	0,000635603	rs3752630	COSV61392290/COSM3998670
TMEM30B	1	14	61747114	A	C	missense_variant	gTg/gGg|p.Val251Gly	0,0104	-	Damaging	0,86174	0,994711518	-	-
TMEM30B	3	14	61747594	G	A	missense_variant	cCc/cTc|p.Pro91Leu	0,0156	0,0398	Tolerated	0,00041	0,003888279	rs137950125	-
TMEM30B	7	14	61748487	C	G	5_prime_UTR_premature_start_codon_gain_variant	|c.-622G>C	0,0364	0,0219	-	-	-	rs77080549	-

Table 2. Percentage of observed chromosomal changes: gains, losses and LOH in TMEM loci in ccRCC tumors (n = 50)

Gene name	Chromosomal localization	Loss		Gain		LOH		No relation to chromosomal aberrations	
		%	n	%	n	%	n	%	n
TMEM213	7q34	0%	0	12%	6	2%	1	86%	43
TMEM207	3q28	4%	2	8%	4	6%	3	82%	41
TMEM30B	14q23.1	28%	14	2%	1	0%	0	70%	35
TMEM72	10q11.21	4%	2	2%	1	2%	1	92%	46
TMEM116	12q24.13	0%	0	4%	2	4%	2	92%	46

Standard settings for SNP arrays in Nexus were adjusted: a cutoff value of 0.15 (> 10 probes), homozygous frequency threshold of 0.95 and minimum loss of heterozygosity (region with LOH) length of 2000 kb and > 50 probes were set. QC measurement in Nexus was used as a measure of the array profile quality (QC < 0.13).

may suggest chromosomal aberrations could partially contribute to its down-regulation in this tumor type.

Tumor microenvironment

It has been previously reported that among all the TCGA analyzed tumors ccRCC has one of the highest levels of the immune infiltration [4]. Since membrane proteins often perform functions related to the immune system, such as receptors of co-receptors, we asked if TMEMs could be implicated in the attraction of immune cells into tumor environment. To test this hypothesis, we performed a gene set enrichment analysis using DEG from the TCGA KIRC RNA-Seq. We observed that the immune infiltration of ccRCC is significantly higher than the presence of immune cells in controls (see **Figure 7**). TREGs were particularly interesting, because they can inhibit immune response, promote cytotoxicity, accelerate T-cell exhaustion, can inhibit the checkpoint receptors, and suppress the secretion of immunosuppressive cytokines [47].

To determine the potential relation of *TMEMs* to the immune score, correlation and partial correlation were calculated, and the genes with correlation/partial correlation of at least +/- 0.3 and FDR lower than 0.01 were selected. Among those genes, *TMEM72* was positively correlated with the TREG signature (0.37, FDR = 4.2826E-19) (**Figure 7B**). *TMEM72* was also partially correlated with the TREG signature (0.36, FDR = 8.7924E-62) (**Figure 7C**). This result suggested that *TMEM72* might potentially play part in the attraction of regulatory T-cells to the tumor microenvironment.

Discussion

Integral membrane proteins are involved in essential cell functions such as bioenergy transduction, transmembrane transfer of nutrients and drugs, signal detection and cell-to-cell communication, adhesion etc. Membrane proteins constitute about 30% of the entire protein content of the cell and comprise about 60% of all pharmaceutical targets and differential regulation of transmembrane proteins could be observed in many cancers, including ccRCC [6, 7].

Structurally TMEM213, 207, 116, 72 and 30B are poorly characterized. Firstly, their predicted membrane bound status has not been confirmed experimentally up to date. Several studies show the proteins localize in membrane-associated environment (TMEM207, 116, 72, 30B), but do not provide the evidence of such association [20, 23, 24, 48]. Here, based on their intracellular dynamics, we confirm that both TMEM213 isoforms, TMEM207, three TMEM116 isoforms, TMEM72_OFR1 and TMEM30B are membrane -bound, except for TMEM72_ORF2 in agreement with predicted loss of transmembrane domains due to alternative splicing. Similarly, the orientation of selected TMEMs within the membrane has not been investigated. Using fluorescence protein protection assay we show that N-terminus of TMEM213, C-termini of TMEM207, 116, and TMEM72 are directed toward the cytoplasm, and both ends of TMEM30B face the cytoplasm.

We also studied the presence of signal peptides on N-termini of these proteins predicted in silico for four TMEMs, with exclusion of

A

	Immunoscore control samples	Immunoscore ccRCC	p-value
B cells (naïve)	0.1165	0.0446	7.75E-23
Macrophages M1	0.032	0.0699	2.30E-16
T cells CD8	0.033	0.1339	3.78E-15
NK cells resting	0.0079	0.0364	1.65E-11
T cells CD4 memory (resting)	0.3083	0.2356	3.48E-11
T cells follicular helper	0.0001	0.0081	3.14E-10
TREGs	**0.003**	**0.0136**	**1.75E-09**
NK cells activated	0.0514	0.0215	2.07E-08
Macrophages M0	0.0033	0.0423	4.02E-08
B cells memory	0.0002	0.0014	0.0063
Dendritic cells activated	0.0059	0.0059	0.0213
T cells memory (activated)	0.0008	0.002	0.0292

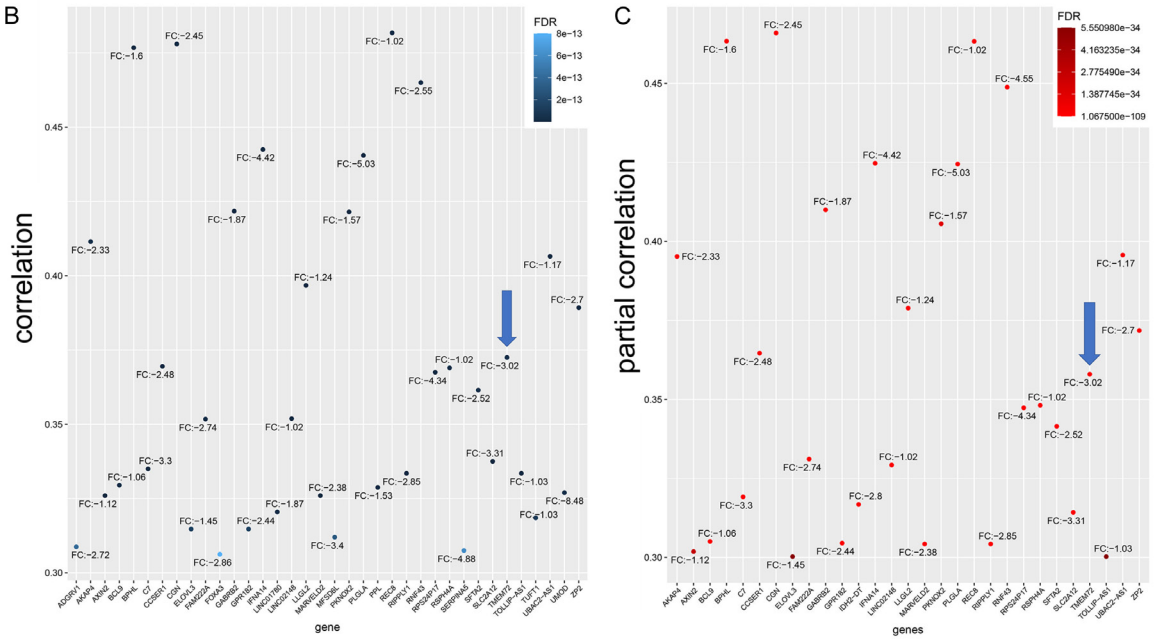


Figure 7. A. Immunoscores obtained from the CIBERSORT analysis. Out of the 22 signatures tested, 12 were significantly different between control and ccRCC tissue (*P*-value < 0.05). B. Positive correlation with Treg immunoscore of downregulated genes. C. Positive partial correlation with Treg immunoscore of downregulated genes.

TMEM30B. The function of these signal peptides direct newly synthesized proteins from the cytoplasm to ER and to end-point organelles. We identified sequences of such signal peptides for TMEM213, 207, and TMEM72_ORF2 and demonstrated localization of expressed constructs in ER.

There is more literature available concerning the localization of the five TMEMs in the cell. TMEM213_ORF1, predicted to be localized in ER, was shown here to be confined to early endosomes, although the second isoform, lacking one AAG codon, was observed in both early endosomes and cytoplasm.TMEM207, 116 and 30B were localized in ER. TMEM30B was previously described to localize in ER in THP-1 and U2OS cells and our observations are in line with these findings [24]. Expression of TMEM116 was found in lung epithelium, but not specified to an organelle [21], while TMEM207 colocalized with WWOX in the cytoplasm in oral squa-

mous cell carcinoma cell lines, and in the cytoplasm of gastric cancer tissues [48]. Just recently, Ding *et al*. published an extensive characterization of the structure of TMEM72, reported TMEM72 likely contains four transmembrane domains and its C-terminus drives anterograde protein transport. The authors found the protein is localized in plasma membrane of HEK293 cells [22]. Here we identified TMEM72 in early endosomes in addition to plasma membrane, which could be likely explained by dynamics of cellular protein transport.

In general, information concerning the function of TMEM213 is limited. Published articles suggest correlation of mRNA expression levels with different patient survival rates as reported for lung adenocarcinoma and head and neck squamous cell carcinoma [17, 18]. Apart from genes grouped into rather general terms, overexpression of TMEM213 leads to de-regulation of genes classified in pathways of ion and transmembrane transport, suggestive of close functional relationship with other membrane proteins.

The involvement of TMEM207 in carcinogenesis was first proposed in gastric signet-ring cell carcinoma and oral squamous cell carcinoma [20, 48]. TMEM207 appears to block the tumor suppressor function of WWOX and may compete with WWOX-interacting oncogenic molecules to impede its tumor suppressor function. Bunai *et al*. found that TMEM207 promoted aerobic glycolysis of oral squamous cell carcinoma by inhibition of WWOX regulation of HIF1a protein. TMEM207 may also participate in intestinal innate immunity since its expression correlated inversely with expression of intelectin-1 [21] and recently combination of Clptm1L and TMEM207 expression was proposed as a prognostic marker in oral squamous cell carcinoma [19]. Interestingly in this study TMEM207 overexpression leads to de-regulation of genes involved in response to hypoxia, although not HIF1a directly, but transcriptional regulator in adaptive response to low oxygen: Hypoxia Inducible Factor 3 Subunit Alpha (HIF3a), which may support its involvement in response to hypoxia.

In contrast to ccRCC tumors full length TMEM116 is highly expressed in non -small-cell lung cancer (NSCLC) tissues and cell lines [49]. In vitro studies suggest that inactivation of TMEM116 inhibits cell proliferation, migration, and invasiveness of human cancer cells. In addition, TMEM116 deficiency inhibited PDK1-AKT-FOXO3A signaling pathway. Our results suggest that overexpression of TMEM116 triggers de-regulation of genes involved in the development of nervous system (presynapse/synapse assembly, regulation of dopamine secretion and synaptic transmission), but also genes designated to negative regulation of angiogenesis pathway. None of the genes encoding proteins in PDK1-AKT-FOXO3A pathway were significantly de-regulated by TMEM116 overexpression [22], but this observation is not surprising since previous reports described experiments on protein, and not mRNA level.

The most interesting pathway enrichment was observed for TMEM72 overexpression, where we find deregulated genes cluster in a group assigned as protein phosphorylation pathway. We know from Ding *et al*. [22], that TMEM72 takes part in the anterograde protein transport in the cell, which might be connected with signal transduction since MAPK13, MAPK15, MAPK4, TGFB2, TLR3, TLR6, TNF exemplify genes categorized to this pathway.

Information on TMEM30B (CDC50B) is also limited. Most of the available data concerns its family member CDC50A - crucial for the activity of multiple P4-ATPases implemented in regulation of proinflammatory cytokines, MAP kinase signaling and NF-κB activation [23]. Van der Velden reported co-immunoprecipitation of TMEM30B with ATP8B1 and ATP8B2, distinctly linking it to P4-ATPases [24]. TMEM30B expression was also reported to correlate with worse overall survival in patients with high-grade serous ovarian cancers [25]. Genes de-regulated by TMEM30B overexpression are grouped into pathways of transcription regulation and biosynthesis, mainly transcription factors, with a significant overlap in genes between the three groups (ZNF, IRF, SOX, HOX genes). We also find ATP2B3 gene in the list of de-regulated genes, supporting previous link between TMEM30B and ATPases.

Finally, we were especially interested in the contribution of TMEMs to ccRCC etiology and the origin of such significant down-regulation of these genes on mRNA level in ccRCC tumors.

Since selected TMEMs are expressed in multiple isoforms we examined exon inclusions and exclusions occurring in ccRCC tissues and found statistically significant exclusion of exon 3 and 10 in TMEM116, suggestive of preferential TMEM116 splicing in this tumor type. In line with ccRCC TCGA data we found a very limited number of mutations in TMEM genes, with one novel and potentially harmful A/C substitution in TMEM30B causing replacement of valine by glycine in the helical part of exoplasmic loop. This observation supports the notion that mutations in TMEM genes are not the cause of their de-regulation in ccRCC. Similar conclusions can be drawn due to the lack of chromosomal aberrations in these genes in ccRCC, with exception of TMEM30B locus, with losses observed in nearly 30% of tumors. These findings point towards a different mechanism of TMEM down-regulation in this kidney cancer. In attempt to find more indications concerning TMEM contribution to cRCC etiology we also performed comparison of gene names between ccRCC TCGA and overexpression datasets (also taking into consideration the directionality of expression changes). Gene ontology analysis of common genes showed gene enrichment in pathways established to be de-regulated in ccRCC tumors, like glycosylation and negative regulation of vasculature development/blood vessel morphogenesis (TMEM213), positive regulation of adaptive immune response (TMEM207), insulin-like growth factor receptor signaling pathway, negative regulation of hormone metabolic process and protein modification (TMEM116). In case of TMEM72, the observation that both datasets shared gene enrichment in protein phosphorylation pathway is especially promising, since it may support the potential involvement of TMEM72 in cell signaling as also indirectly suggested by our in-silico analysis of tumor microenvironment. Regulation of transcription and biosynthesis showed up as common with TMEM30B. Although indirect, this comparison serves as an indicator of cellular processes the five TMEM proteins may contribute to in ccRCC tumors.

In summary, TMEM213 protein localizes in early endosomes, contains a signal peptide on N-terminus, its C-terminus is directed towards cytoplasm, and it may contribute to processes like glycosylation and regulation of vasculature in ccRCC tumors. We confirmed TMEM207 to be a membrane protein localized in EE, with N-terminal cleavable signal peptide and C-terminus directed towards the cytoplasm, likely involved in cellular biosynthesis and regulation of adaptive immune response in ccRCC. TMEM116 was found to localize in ER with C-terminus facing the cytoplasm and to contain cleavable signal peptide on N-terminus. It may also contribute to ccRCC etiology with its involvement in protein modification processes. We found TMEM72 to localize in early endosomes and plasma membrane, with C-terminus facing the cytoplasm. It also contained N-terminal cleavable signal peptide and may take part in protein phosphorylation/signaling in ccRCC tumors. Lastly, TMEM30B we found TMEM30B to be confined to ER, with both ends being cytoplasmic and likely involved in regulation of transcription and biosynthesis in ccRCC.

All in all, the body of information on TMEM structure, localization and function slowly becomes more extensive and complementary data may soon lead to unraveling their exact function in the cell. Although we obtained vast data on the structure of TMEM213, 207, 116, 72 and 30B, their exact contribution to ccRCC etiology requires validation and further, in-depth studies.

Acknowledgements

We would like to thank Daria Niewiadomska for the technical support. This work was funded by the National Science Centre Poland 2014/15/B/NZ2/00589 (J.W.).

Disclosure of conflict of interest

None.

Address correspondence to: Dr. Joanna Wesoly, Laboratory of High Throughput Technologies, Faculty of Biology, Adam Mickiewicz University, Uniwersytetu Poznanskiego 6, Poznań 61614, Poland. E-mail: j.wesoly@amu.edu.pl

References

[1] American Cancer Society, Kidney Cancer Treatment, https://www.cancer.org/content/dam/CRC/PDF/Public/8662.00.pdf. 2021.

[2] Siegel RL, Miller KD, Fuchs HE and Jemal A. Cancer statistics, 2021. CA Cancer J Clin 2021; 71: 7-33.

[3] Dawson M and Bannister A. Jak-stat signaling: from basics to disease. 2012. pp. 27-46.
[4] Şenbabaoğlu Y, Gejman RS, Winer AG, Liu M, Van Allen EM, de Velasco G, Miao D, Ostrovnaya I, Drill E, Luna A, Weinhold N, Lee W, Manley BJ, Khalil DN, Kaffenberger SD, Chen Y, Danilova L, Voss MH, Coleman JA, Russo P, Reuter VE, Chan TA, Cheng EH, Scheinberg DA, Li MO, Choueiri TK, Hsieh JJ, Sander C and Hakimi AA. Tumor immune microenvironment characterization in clear cell renal cell carcinoma identifies prognostic and immunotherapeutically relevant messenger RNA signatures. Genome Biol 2016; 17: 231.
[5] Cancer Genome Atlas Research Network. Comprehensive molecular characterization of clear cell renal cell carcinoma. Nature 2013; 499: 43-49.
[6] Wrzesiński T, Szelag M, Cieślikowski WA, Ida A, Giles R, Zodro E, Szumska J, Poźniak J, Kwias Z, Bluyssen HA and Wesoly J. Expression of pre-selected TMEMs with predicted ER localization as potential classifiers of ccRCC tumors. BMC Cancer 2015; 15: 518.
[7] Schmit K and Michiels C. TMEM proteins in cancer: a review. Front Pharmacol 2018; 9: 1345.
[8] Cuajungco MP, Podevin W, Valluri VK, Bui Q, Nguyen VH and Taylor K. Abnormal accumulation of human transmembrane (TMEM)-176A and 176B proteins is associated with cancer pathology. Acta Histochem 2012; 114: 705-712.
[9] Hrašovec S, Hauptman N, Glavač D, Jelenc F and Ravnik-Glavač M. TMEM25 is a candidate biomarker methylated and down-regulated in colorectal cancer. Dis Markers 2013; 34: 93-104.
[10] Pérez-Magán E, Campos-Martín Y, Mur P, Fiaño C, Ribalta T, García JF, Rey JA, Rodríguez de Lope A, Mollejo M and Meléndez B. Genetic alterations associated with progression and recurrence in meningiomas. J Neuropathol Exp Neurol 2012; 71: 882-893.
[11] Abermil N, Guillaud-Bataille M, Burnichon N, Venisse A, Manivet P, Guignat L, Drui D, Chupin M, Josseaume C, Affres H, Plouin PF, Bertherat J, Jeunemaître X and Gimenez-Roqueplo AP. TMEM127 screening in a large cohort of patients with pheochromocytoma and/or paraganglioma. J Clin Endocrinol Metab 2012; 97: E805-809.
[12] Fuller CM. Time for TMEM? J Physiol 2012; 590: 5931-5932.
[13] Ma F, Zhang C, Prasad KV, Freeman GJ and Schlossman SF. Molecular cloning of Porimin, a novel cell surface receptor mediating oncotic cell death. Proc Natl Acad Sci U S A 2001; 98: 9778-9783.
[14] Foulquier F, Amyere M, Jaeken J, Zeevaert R, Schollen E, Race V, Bammens R, Morelle W, Rosnoblet C, Legrand D, Demaegd D, Buist N, Cheillan D, Guffon N, Morsomme P, Annaert W, Freeze HH, Van Schaftingen E, Vikkula M and Matthijs G. TMEM165 deficiency causes a congenital disorder of glycosylation. Am J Hum Genet 2012; 91: 15-26.
[15] Carette JE, Guimaraes CP, Varadarajan M, Park AS, Wuethrich I, Godarova A, Kotecki M, Cochran BH, Spooner E, Ploegh HL and Brummelkamp TR. Haploid genetic screens in human cells identify host factors used by pathogens. Science 2009; 326: 1231-1235.
[16] Ishikawa H and Barber GN. STING is an endoplasmic reticulum adaptor that facilitates innate immune signalling. Nature 2008; 455: 674-678.
[17] Zou J, Li Z, Deng H, Hao J, Ding R and Zhao M. TMEM213 as a novel prognostic and predictive biomarker for patients with lung adenocarcinoma after curative resection: a study based on bioinformatics analysis. J Thorac Dis 2019; 11: 3399-3410.
[18] Koteluk O, Bielicka A, Lemańska Ż, Jóźwiak K, Klawiter W, Mackiewicz A, Kazimierczak U and Kolenda T. The landscape of transmembrane protein family members in head and neck cancers: their biological role and diagnostic utility. Cancers (Basel) 2021; 13: 4737.
[19] Hano K, Hatano K, Saigo C, Kito Y, Shibata T and Takeuchi T. Combination of Clptm1L and TMEM207 expression as a robust prognostic marker in oral squamous cell carcinoma. Front Oral Health 2021; 2: 638213.
[20] Bunai K, Okubo H, Hano K, Inoue K, Kito Y, Saigo C, Shibata T and Takeuchi T. TMEM207 hinders the tumour suppressor function of WWOX in oral squamous cell carcinoma. J Cell Mol Med 2018; 22: 1026-1033.
[21] Maeda K, Saigo C, Kito Y, Sakuratani T, Yoshida K and Takeuchi T. Expression of TMEM207 in colorectal cancer: relation between TMEM207 and Intelectin-1. J Cancer 2016; 7: 207-213.
[22] Ding J, Matsumiya T, Miki Y, Hayakari R, Shiba Y, Kawaguchi S, Seya K and Imaizumi T. ER export signals mediate plasma membrane localization of transmembrane protein TMEM72. FEBS J 2022; [Epub ahead of print].
[23] van der Mark VA, Ghiboub M, Marsman C, Zhao J, van Dijk R, Hiralall JK, Ho-Mok KS, Castricum Z, de Jonge WJ, Oude Elferink RP and Paulusma CC. Phospholipid flippases attenuate LPS-induced TLR4 signaling by mediating endocytic retrieval of Toll-like receptor 4. Cell Mol Life Sci 2017; 74: 715-730.
[24] van der Velden LM, Wichers CG, van Breevoort AE, Coleman JA, Molday RS, Berger R, Klomp

LW and van de Graaf SF. Heteromeric interactions required for abundance and subcellular localization of human CDC50 proteins and class 1 P4-ATPases. J Biol Chem 2010; 285: 40088-40096.
[25] Li Y, Jaiswal SK, Kaur R, Alsaadi D, Liang X, Drews F, DeLoia JA, Krivak T, Petrykowska HM, Gotea V, Welch L and Elnitski L. Differential gene expression identifies a transcriptional regulatory network involving ER-alpha and PITX1 in invasive epithelial ovarian cancer. BMC Cancer 2021; 21: 768.
[26] Babraham Bioinformatics. Available from: https://www.bioinformatics.babraham.ac.uk/index.html.
[27] Ewels P, Magnusson M, Lundin S and Käller M. MultiQC: summarize analysis results for multiple tools and samples in a single report. Bioinformatics 2016; 32: 3047-3048.
[28] Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M and Gingeras TR. STAR: ultrafast universal RNA-seq aligner. Bioinformatics 2013; 29: 15-21.
[29] Li H. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. Bioinformatics 2011; 27: 2987-2993.
[30] Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G and Durbin R. The sequence alignment/map format and SAMtools. Bioinformatics 2009; 25: 2078-2079.
[31] Liao Y, Smyth GK and Shi W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics 2014; 30: 923-930.
[32] Alexa A, Rahnenführer J and Lengauer T. Improved scoring of functional groups from gene expression data by decorrelating GO graph structure. Bioinformatics 2006; 22: 1600-1607.
[33] Deryło K, Michalec-Wawiórka B, Krokowski D, Wawiórka L, Hatzoglou M and Tchórzewski M. The uL10 protein, a component of the ribosomal P-stalk, is released from the ribosome in nucleolar stress. Biochim Biophys Acta Mol Cell Res 2018; 1865: 34-47.
[34] Lorenz H, Hailey DW, Wunder C and Lippincott-Schwartz J. The fluorescence protease protection (FPP) assay to determine protein localization and membrane topology. Nat Protoc 2006; 1: 276-279.
[35] Horton P, Park KJ, Obayashi T, Fujita N, Harada H, Adams-Collier CJ and Nakai K. WoLF PSORT: protein localization predictor. Nucleic Acids Res 2007; 35: W585-587.
[36] Hiller K, Grote A, Scheer M, Münch R and Jahn D. PrediSi: prediction of signal peptides and their cleavage positions. Nucleic Acids Res 2004; 32: W375-379.
[37] Frank K and Sippl MJ. High-performance signal peptide prediction based on sequence alignment techniques. Bioinformatics 2008; 24: 2172-2176.
[38] Käll L, Krogh A and Sonnhammer EL. Advantages of combined transmembrane topology and signal peptide prediction-the Phobius web server. Nucleic Acids Res 2007; 35: W429-432.
[39] Almagro Armenteros JJ, Salvatore M, Emanuelsson O, Winther O, von Heijne G, Elofsson A and Nielsen H. Detecting sequence signals in targeting peptides using deep learning. Life Sci Alliance 2019; 2: e201900429.
[40] Camacho Londoño J and Philipp SE. A reliable method for quantification of splice variants using RT-qPCR. BMC Mol Biol 2016; 17: 8.
[41] Andrews S. FastQC: a quality control tool for high throughput sequence data. 2010.
[42] Chen S, Zhou Y, Chen Y and Gu J. fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics 2018; 34: i884-i890.
[43] Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv: Genomics 2013.
[44] Sandmann S, Karimi M, de Graaf AO, Rohde C, Göllner S, Varghese J, Ernsting J, Walldin G, van der Reijden BA, Müller-Tidow C, Malcovati L, Hellström-Lindberg E, Jansen JH and Dugas M. appreci8: a pipeline for precise variant calling integrating 8 tools. Bioinformatics 2018; 34: 4205-4212.
[45] Newman AM, Liu CL, Green MR, Gentles AJ, Feng W, Xu Y, Hoang CD, Diehn M and Alizadeh AA. Robust enumeration of cell subsets from tissue expression profiles. Nat Methods 2015; 12: 453-457.
[46] Kluzek K, Srebniak MI, Majer W, Ida A, Milecki T, Huminska K, van der Helm RM, Silesian A, Wrzesinski TM, Wojciechowicz J, Beverloo BH, Kwias Z, Bluyssen HAR and Wesoly J. Genetic characterization of Polish ccRCC patients: somatic mutation analysis of PBRM1, BAP1 and KDMC5, genomic SNP array analysis in tumor biopsy and preliminary results of chromosome aberrations analysis in plasma cell free DNA. Oncotarget 2017; 8: 28558-28574.
[47] Togashi Y, Shitara K and Nishikawa H. Regulatory T cells in cancer immunosuppression-implications for anticancer therapy. Nat Rev Clin Oncol 2019; 16: 356-371.
[48] Takeuchi T, Adachi Y and Nagayama T. A WWOX-binding molecule, transmembrane protein 207, is related to the invasiveness of gastric signet-ring cell carcinoma. Carcinogenesis 2012; 33: 548-554.
[49] Zhang S, Dai H, Li W, Wang R, Wu H, Shen M, Hu Y, Xie L and Xing Y. TMEM116 is required for lung cancer cell motility and metastasis through PDK1 signaling pathway. Cell Death Dis 2021; 12: 1086.