3.3 Subcellular localization of Hsa21 proteins

At a time point when 96 Hsa21 full ORFs had been cloned and sequence-verified, this set was chosen for use in transfected cell array experiments to perform a rapid and cost-effective analysis of the subcellular localization of Hsa21 proteins. The project was performed in collaboration with the cell array group lead by Dr. Michal Janitz at the Max Planck Institute for Molecular Genetics, Berlin. This collaboration resulted in a joint publication (Hu *et al.* 2006).

Briefly, ORFs were cloned into a mammalian expression vector encoding an aminoterminal hexahistidine tag. Insert-verified plasmids were purified and diluted with gelatin solution. Prior to spotting, microscope glass slides were treated with poly-L-lysine for better retention of DNA and cells. Plasmid-gelatin solutions were spotted manually or by an array-printer. HEK293T cells were grown on the slides, reverse transfected with the constructs and cultured in medium for ≤72 hours. Hsa21 proteins expressed on the cell arrays were detected using anti-hexahistidine antibody, and additional immunolocalization assays were applied for cell organelle colocalizations. Protein expression and labeling was monitored using a BioCCD laser scanning system with different filters. For single-cell analysis, fluorescent and confocal microscopes were used.

Subcloning of the Hsa21 ORFs into mammalian expression vectors and sequence verifications were performed by the author of this dissertation (H.-J.W.). Cell array experiments were performed by a graduate student in the collaborating group (Yuhui Hu, Y.H.H.). Data analysis was performed together by both Y.H.H and H.-J.W. The optimization and application of protein localization detection on cell arrays is described in the dissertation of Y.H.H., which is available online at http://www.diss.fu-berlin.de/2006/577/.

In the following chapters, the cell array experiments and results will be described in more detail.

3.3.1 Cloning of mammalian expression constructs

Constructs for expression with N-terminal $(His)_6$ tag

A total of 96 cloned Hsa21 ORFs were selected for determination of subcellular localizations using transfected cell arrays. Using LR clonase recombination reaction,

all ORFs were transferred from the sequence-verified pENTR201-ORF entry clones into the Gateway-compatible mammalian expression vector pDEST26 (Invitrogen). This vector drives protein expression with amino-terminal $(His)_6$ fusion tag through a constitutive CMV promoter (Figure 3-7A). After transformation of the ORF constructs into *E. coli* DH5α, insert sizes were verified by colony PCR using the following primers:

```
pEXP26_insert_forward: 5' - CGG ACC ATG GCG TAC TAC C - 3'
pEXP26_insert_reverse: 5' - TAA TAC GAC TCA CTA TAG GG - 3'
```

This procedure resulted in 89 constructs with correct insert size for localization experiments on transfected cell arrays (93% success rate).

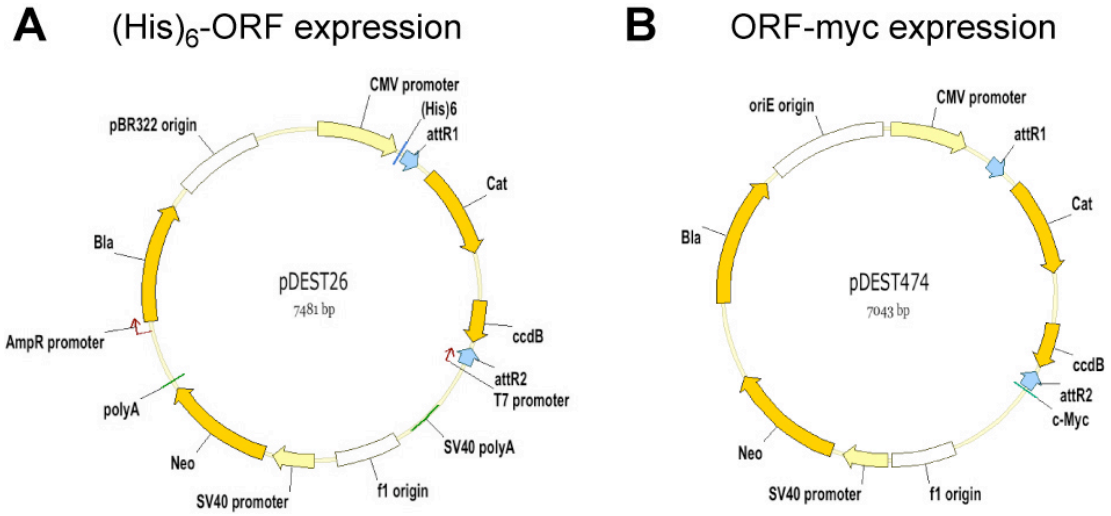

Figure 3-7. Vector maps of pDEST26 and pDEST474 used for transient fusion protein expression in reverse transfected HEK293T cells. A: pDEST26 was used for expression with N-terminal hexahistidine fusion tag. B: pDEST474 was used for expression with C-terminal myc fusion tag. Abbreviations: CMV – constitutive cytomegalovirus immediate early promoter; (His)6 – hexahistidine fusion tag; attR1/2 – Gateway recombination sites for integration of ORF; Cat – Chloramphenicol acetyltransferase used for vector propagation in *E. coli* DB3.1 strain; ccdB – controller of cell division or death B gene, inhibits bacterial gyrase; T7 promoter – used for colony PCR and sequencing of insert; c-Myc – c-myc epitope fusion tag; SV40 polyA – Simian virus 40 polyadenylation signal; f1 origin – phage f1 origin of replication, can be used for production of single-stranded DNA; SV40 promoter – constitutive simian virus 40 early promoter; Neo – Aminoglycoside phosphotransferase, can be used for G418 selection of stably transfected cells; polyA – synthetic polyadenylation signal; AmpR promoter – constitutive *E. coli* promoter; Bla – β-lactamase, used for Ampicillin selection and propagation of ORF-containing constructs (pEXP26-/pEXP474-ORFs) in *E. coli*; pBR322 origin – origin of replication derived from plasmid pBR322; oriE origin – origin of replication from *E. coli*.

Constructs for expression with C-terminal c-myc tag

In order to confirm the patterns of protein localizations obtained by N-terminal tagging, constructs were prepared encoding carboxy-terminally tagged proteins. Seventeen Hsa21 ORFs – localizing to a variety of different cellular compartments –

were PCR-amplified from pENTR201-ORF entry clones with primer pairs designed to include a Kozak consensus sequence around the start codon (ACC ATG …) and to remove the stop codon. Additional information on the primers used for subcloning can be found in the Appendix, chapter 6.2.2 ('Primer for C-terminal epitope tagging').

The resulting PCR products were cloned into Gateway donor vector pDONR/Zeo (Invitrogen) and verified by 5' and 3' insert sequencing. Entry vectors with correct insert were used to transfer the ORFs into the Gateway-compatible mammalian expression vector pDEST474 (kind gift of Dr. Esposito, NCI-Frederick Vector Engineering Group). This vector drives protein expression with carboxy-terminal c-myc fusion tag through a constitutive CMV promoter (Figure 3-7B). After transformation of the ORF constructs into *E. coli* DH5α, insert sizes were verified by colony PCR using the following primers:

```
pEXP474_insert_forward: 5' – GCC TGC AGG TAC CGG ATC A – 3'
pEXP474_insert_reverse: 5' – G CTT TTG TTC GCT AGC CAT TC – 3'
```

This procedure resulted in 17 constructs with correct insert size for localization experiments on transfected cell arrays (100% cloning success rate).

3.3.2 Antibody selection for hexahistidine and cell organelles

The success of immunocytochemical protein detection greatly depends on the set-up and the quality of the antibodies used. Direct immunodetection with fluorochrome-conjugated primary antibodies is time-efficient, while indirect immunodetection using primary antibodies plus fluorochrome-conjugated secondary antibodies can yield better signal amplification.

Selection of antibodies for hexahistidine detection

To test the performance of antibodies from different manufactures to recognize the hexahistidine tag, $(His)_6$-tagged eGFP (enhanced GFP) was expressed on transfected cell arrays. Then, a number of commercially available anti-His antibodies was used to label the $(His)_6$-eGFP protein – either directly or indirectly – with red fluorophores using different secondary antibodies. The labeling performance of all antibody combinations was evaluated based on the overlap between eGFP signal and red signal. In total, eight primary antibodies were screened, and most antibodies resulted in unspecific staining or low signal-to-noise ratio, while the direct Penta•His-

Alexa555 and Penta•His-Alexa488 antibodies (Qiagen) gave a sensitive and specific labeling (for further details, see the dissertation of Y.H.H.).

Selection of stains for cell organelles

For a protein localization study, it is essential to unambiguously identify subcellular compartments. In the study described here, two approaches (control proteins and organelles counterstaining) were applied to test the applicability of transfected cell arrays for the determination of subcellular localizations. First, expression vectors for a number of control proteins with known localization were constructed following the established cloning protocol and used for reverse transfection and immunostaining. These control proteins (KDELR1, LMNA, PEX11A, TGN38, CDH8, and GFP) were found to localize in the predicted compartments, proving the general applicability of the cell arrays for this purpose. The localization images can be accessed online at http://www.molgen.mpg.de/~ag_onf/chr21ORF-SubLoc/index.html.

Then, a high-throughput colocalization approach on cell arrays was developed for the different cellular compartments. Organelle-specific antibodies or dyes from different suppliers were screened, and the protocols for their application were optimized. In total, nine reagents were found to provide sensitive and specific organelle labeling (Table 3-3). These organelle stains were subsequently used for counterstaining with the Penta•His AlexaFluor antibodies for colocalization studies with Hsa21 proteins.

Table 3-3. Reagents for organelle counterstains used in colocalization studies on cell arrays.

Compartment	Reagent	Supplier
Nucleus	DAPI stain	Sigma-Aldrich
ER	Mouse anti-PDI IgG	Stressgen
Golgi	Mouse anti-Adaptin-γ IgG	BD Biosciences
Lysosomes	Mouse anti-LAMP2 IgG	H4B4, Developmental Studies Hybridoma Bank, USA
Mitochondria	Rabbit anti-Prohibitin IgG	Abcam
Peroxisomes	Rabbit anti-Catalase IgG	Abcam
F-actin	Rhodamine-conjugated Phalloidin	Molecular Probes
Microtubules	Mouse anti-α-tubulin IgG	Sigma-Aldrich
Intermediate filaments	Mouse anti-Vimentin	Affinity BioReagents

Modified from Table 4.2 (Dissertation of Y.H.H., available online at http://www.diss.fu-berlin.de/2006/577/).

3.3.3 Determination of subcellular localizations on cell arrays

Preparation of plasmid DNA and reverse transfection

To prepare gelatin-DNA samples for array spotting, purified plasmid DNAs were diluted 10-fold with 0.2% gelatin solution to a final concentration of 0.04-0.07 µg/µl. After dilution, the gelatin-DNA samples were mixed and stored at 4°C before printing.

Subcellular localization of Hsa21 proteins

All 89 Hsa21 full ORFs that had been subcloned and verified as pEXP26-ORFs were overexpressed on HEK293T cell arrays and detected with Penta-His AlexaFluor antibodies. Additional information on this cell line can be found in the Appendix, chapter 6.3.4 ('*Homo sapiens* cell line').

In total, 52 out of 89 proteins (58%) could be detected on the cell arrays, while the other tested proteins were expressed at a level below the detection threshold of this method. The subcellular localization patterns of the 52 detected proteins were determined through counterstaining with several cell compartment markers, as described above. As result, the subcellular localization of 28 Hsa21 proteins was described for the first time. All obtained localization data is summarized in Table 3-4, together with protein accession numbers, gene functions and previously described localization patterns. The localization images not shown here can be accessed online at http://www.molgen.mpg.de/~ag_onf/chr21ORF-SubLoc/index.html.

Table 3-4. Experimentally determined subcellular localizations of 52 Hsa21 proteins.

Gene Symbol	Gene Function	Localization in HEK293T[a]	Localization (SwissProt)[a]	OMIM Identifier[b]	Protein Accession
ABCG1	cholesterol transporter	Golgi / PM	ER / Golgi	603076	CAA62631
AGPAT3	acyltransferase	ER / PM	—		AAH11971
B3GALT5	galactosyl-transferase	ER / Golgi	Golgi	604066	NP_006048
BACH1	transcriptional regulator	Nuc (M-phase) / Cyto (punct.)	Nuc (predict.)	602751	BAA24932
C21orf4	unknown	PM	—		AAC05974
C21orf7 (TAK1L)	unknown	Nuc / Cyto	—		AAF81754
C21orf19	unknown	Nuc / Cyto	—		AAL34462

Gene Symbol	Gene Function	Localization in HEK293T[a]	Localization (SwissProt)[a]	OMIM Identifier[b]	Protein Accession
C21orf25	unknown	Nuc / Cyto	—		XP_032945
C21orf30	unknown	Nuc	—		CAB56001
C21orf59	unknown	Nuc / Cyto	—		AAG00496
C21orf66 (GCFC)	DNA-binding factor candidate	Cyto	Nuc (predict.)		AAD34617
C21orf69	unknown	ER	—		AAK60445
C21orf96	unknown	Cyto (punct.)	—		NP_079419
C21orf103 (KRTAP6-1)	unclear	Cyto	—		NP_853633
CBS	cystathionine beta-synthase	Cyto	Cyto	236200 ALLELIC VARIANTS / CLINICAL SYNOPSIS	NP_000062 (splicing isoform)
CCT8	chaperonin	Cyto	Cyto		NP_006576
CHAF1B	chromatin assembly factor	Nucleoplasm (interphase) / Cyto (M phase)	Nuc / Cyto (M phase)	601245	NP_005432
CLDN8	tight junction	ER / PM	—		NP_955360
CLDN14	tight junction	ER / PM	—	605608 ALLELIC VARIANTS	AAG60052
CLDN17	tight junction	Golgi / PM	—		CAB60616
CRYZL1	unknown	Cyto	—	603920	BAA91605
CXADR	plasma membrane receptor	PM	apical junction / basolateral membrane	602621	NP_001329
DNMT3L	DNA methylation	Nuc / Cyto	Nuc (predict.)	606588	AAH02560
DSCR3	unknown	Nuc	—	605298	NP_006043
ETS2	transcriptional regulator	Nuc	Nuc (predict.)	164740	NP_005230
HLCS	holocarboxylase synthetase	Cyto	Cyto / Mito	253270 CLINICAL SYNOPSIS	NP_000402
HMGN1	DNA binding	Nucleoplasm	Nuc	163920	AAA52676
HSF2BP	transcription factor binding	Cyto	—	604554	NP_008962
IFNGR2	interferon receptor	ER / PM	—	209950 CLINICAL SYNOPSIS	AAH03624
KCNE1	K^+ channel	PM / Lyso	—	176261 ALLELIC VARIANTS / CLINICAL SYNOPSIS	AAH36452
KCNE2	K^+ channel	PM / Lyso	—	603796 ALLELIC VARIANTS	NP_005127

Gene Symbol	Gene Function	Localization in HEK293T[a]	Localization (SwissProt)[a]	OMIM Identifier[b]	Protein Accession
KCNJ6	K^+ channel	Golgi / PM	—	600877	NP_002231
KCNJ15	K^+ channel	Golgi / PM	—	602106	NP_002234
KIAA0179 (RRP1B)	rRNA processing	Nuc / Cyto (punct.)	Nuc (Nucleoli)		BAA11496
MCM3AP	DNA replication inhibitor	Nuc / Cyto	—	603294	BAA25170
MxA (MX1)	GTP-binding protein	Cyto (punct.)	Cyto	147150	NP_002453
PCBP3	RNA binding	Nuc / Cyto	Cyto	608502	AAH12061
PCP4	calmodulin regulator	Nuc / Cyto	—	601629	CAA63724
PDE9A	phospho-diesterase	Cyto (accum.)	—	602973	AAH09047
PDXK	pyridoxal kinase	Cyto	Cyto	179020	AAH00123
PFKL	phosphofructo-kinase	Cyto (accum.)	—	171860	AAH09919
PKNOX1	transcriptional regulator	Nuc / Cyto	Nuc (predict.)	602100	AAH07746
PPIAL3	peptidylprolyl isomerase A-like 3	Nuc / Cyto	—	123840	CAA37039
RPS5L	ribosomal protein S5-like	Cyto	—		98% identity to BAB79493
RRP1 (NNP1)	rRNA processing	Nucleoli	Nuc (Nucleoli)		AAH00380
SH3BGR	SH3 adapter protein	Cyto	—	602230	AAH06371
TMPRSS3a	transmembrane protease	ER	ER	605511 ALLELIC VARIANTS	NP_076927
TSGA2	chromosome associated	Nuc / Cyto	Cyto / Nuc (meiosis) (predict.)		NP_543136
UBASH3A	endocytosis control	Cyto	Nuc (predict.)	605736	NP_061834
UBE2G2	ubiquitin-conjugating enzyme	Cyto	—	603124	AAC32312
WDR4	tRNA modification	Nucleoplasm	—	605924	AAH06341
WDR9[c] *(BRWD1_3')*	unknown	Nuc	—		BAD74072

Localization patterns were generally consistent with described gene functions. [a] accum. – accumulated; Cyto – cytoplasm; ER – endoplasmic reticulum; Lyso – lysosomes and endosomes; Nuc: nucleus; PM: plasma membrane; predict. – three non-experimental localization qualifiers including “potential”, “probable” and “by similarity”; punct. – punctuate; [b] 33 genes had an OMIM entry, among which seven had known allelic variants associated with diseases, and four had a clinical synopsis. [c] BRWD1_3' stands for a 3' fragment of the BRWD1 (WDR9) ORF, containing nucleotides 5663..6949 of RefSeq mRNA NM_033656. Modified from Table 1 (Hu *et al.* 2006).

A large fraction of the detected Hsa21 proteins (71%) was found either in the cytosol (31%), in the nucleus (17%) or in both compartments (23%), sometimes with transport observed between the two localizations. The other 29% of the proteins were associated with components of the secretory pathway, including the endoplasmic reticulum (10%), the Golgi complex (2%), the plasma membrane (13%) and the endosome/lysosome compartment (4%). No protein was found localized to mitochondria, peroxisomes or cytoskeletal structures. Examples of localizations to different cellular compartments are shown in Figure 3-8. In the following paragraphs, some examples regarding protein distribution to different compartments will be described in more detail.

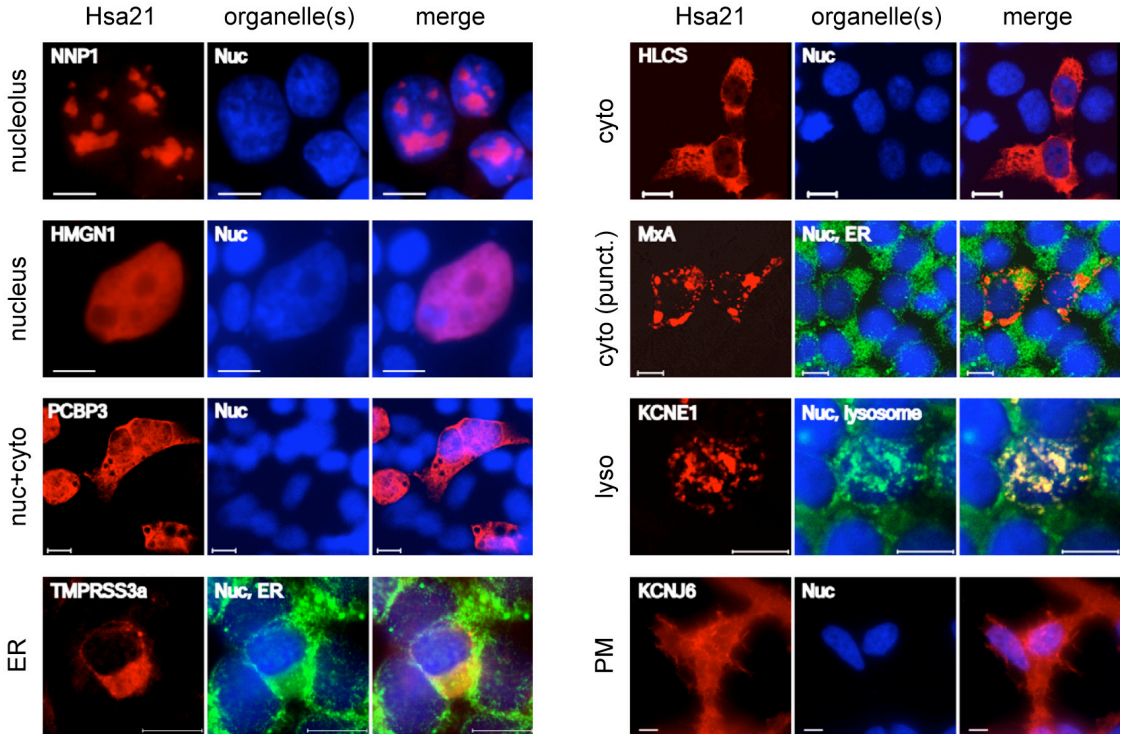

Figure 3-8. Localization categories for detected Hsa21 proteins. The Hsa21 proteins were localized in a variety of cellular compartments. Images in the 'Hsa21' panels represent anti-hexahistidine staining of Hsa21 proteins, with the gene symbols indicated in the upper left corner of each image. Images in the 'organelle(s)' panel show counterstainings, with organelles indicated in the upper left corner. Images in the 'merge' panels represent merged pictures from detection of Hsa21 proteins and counterstainings. Abbreviations: nuc – nucleus; cyto – cytoplasm; cyto (punct.) – punctuate staining in the cytoplasm; lyso – lysosomes; PM – plasma membrane. Scale bars represent 10 µm. Figure modified from Figure 2 (Hu *et al.* 2006).

Nucleus and nucleoli

17% of the Hsa21 proteins were found to localize to the nucleus. However, different localization patterns could be observed inside this compartment. For example, the ribosomal RNA processing protein 1 (RRP1, alias novel nuclear protein 1, NNP1)

localized exclusively to the nucleoli (Figure 3-8; NNP1). In contrast, the non-histone chromosomal protein HMGN1 was found mainly in the nucleoplasm outside of the nucleoli (Figure 3-8; HMGN1). Other proteins seen mainly in the nucleoplasm are CHAF1B and WDR4.

Nucleus and cytoplasm

In total, 23% of the Hsa21 proteins were found to localize to both the nucleus and the cytoplasm. The nucleus/cytoplasm distribution ratio for particular proteins, however, often varied from cell to cell. In some cells, proteins with dual localization could be found either only in the nucleus, or only in the cytoplasm, suggesting a continuous translocation activity. For example, the distribution of the poly(rC)-binding protein 3 (PCBP3) was found to be remarkably versatile, ranging from nucleus-only over both nucleus and cytoplasm to cytoplasm-only (Figure 3-8; PCBP3).

Cytoplasm

Two distinct localization features were observed for the 31% of Hsa21 proteins that localized to the cytoplasm. Proteins were either distributed evenly throughout the entire cytosol (12/16 proteins), or formed punctuate patterns (4/16 proteins). For example, the metabolic enzyme holocarboxylase synthetase (HLCS) was found to localize throughout the cytosol (Figure 3-8; HLCS), whereas the myxovirus resistance protein 1 (MX1/MxA) revealed a punctate pattern in the cytoplasm, which does not overlap with any organelle staining (Figure 3-8; MxA). It has been reported that the MxA protein can localize to the smooth ER and might inhibit viral replication through alterations in membrane organization (Stertz *et al.* 2006). However, no colocalization of MxA with the ER marker protein disulfide isomerase (PDI) was observed here. Other proteins seen in punctuate patterns are C21orf96, BACH1, RRP1B, PDE9A and PFKL.

Endoplasmic reticulum and Golgi apparatus

Altogether, 12% of the Hsa21 proteins were localized to first stations of the secretory pathway, the endoplasmic reticulum (ER) and the Golgi apparatus. Proteins resident only in these compartments are actively retained to escape the default pathway. For example, the transmembrane serine protease TMPRSS3, which is associated with autosomal recessive deafness, was found to localize exclusively to the ER (Figure 3-

8; TMPRSS3a). Other proteins observed mainly in the ER are AGPAT3, B3GALT5 and C21orf69. In contrast to the ER, no Hsa21 protein was found exclusively localized to the Golgi apparatus. The proteins observed in the Golgi were also seen either in the ER (B3GALT5) or in the plasma membrane (ABCG1, CLDN17, KCNJ6 and KCNJ15).

Plasma membrane, endosomes and lysosomes

Of the 17% of Hsa21 proteins that localized to the late stations of the secretory pathway (plasma membrane or endosome/lysosome compartment), only two were found exclusively at the plasma membrane, namely the coxsackie virus and adenovirus receptor (CXADR) and the unknown protein C21orf4. Other plasma membrane proteins, such as the tight junction proteins CLDN8 and CLDN14 as well as the interferon gamma receptor IFNGR2, were also observed in the ER. A number of proteins, such as the voltage-gated potassium channels KCNE1, KCNE2 and KCNJ6, showed varying subcellular localizations, reaching from ER and Golgi over the plasma membrane to the endosome/lysosome compartment (Figure 3-8; KCNJ6 and KCNE1; see also next paragraph).

3.3.4 Protein translocation

While many Hsa21 proteins showed relatively stable subcellular localizations when comparing multiple cells on the arrays, some proteins could be identified that had a variety of distinct localization patterns. Two main routes of protein translocation could be observed on the cell arrays: (1) post-translational protein trafficking along the secretory pathway, and (2) cell cycle dependent protein shuttling between nucleus and cytoplasm.

Protein trafficking along the secretory pathway

For some proteins, the entire secretory pathway could be traced throughout the different organelles. Good examples are two plasma membrane proteins KCNE1 and KCNE2. Figure 3-9A shows consecutive steps of KCNE2 protein maturation, with its final destination being the PM, where it acts as voltage-gated potassium channel (subfamily type E). In this study, it was found for the first time that both proteins colocalized also with the lysosome marker LAMP2 (lysosomal-associated membrane glycoprotein 2). These findings could either reflect the ability of KCNE1 and KCNE2

to shuttle between the PM and the endosome/lysosome compartment, or alternatively, the lysosomal location of these proteins could result from their direct transport from the Golgi to the lysosomes. Interestingly, the potassium channel protein KCNJ6, a member of the subfamily J, never showed the lysosomal pattern and localized at the PM (Figure 3-8), which is in accordance with different biological functions of the respective potassium channel subfamilies E and J.

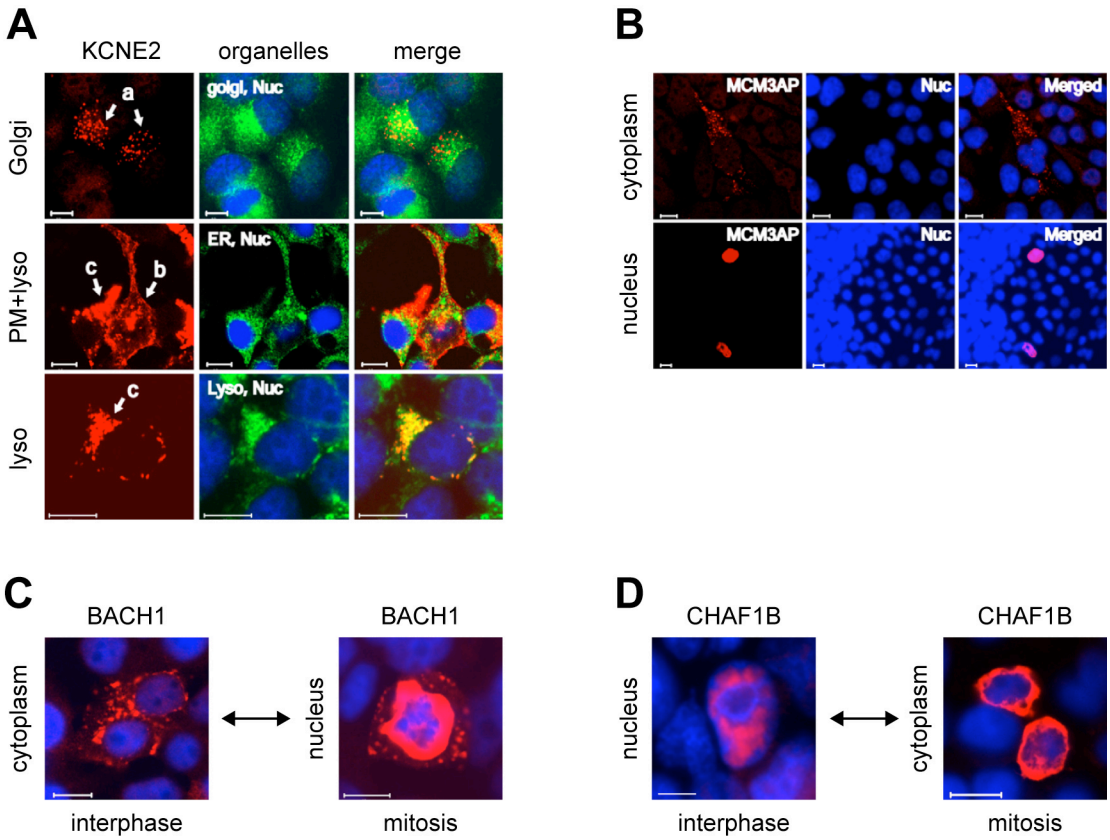

Figure 3-9. Intracellular translocation of Hsa21 proteins observed on cell arrays. A: Stages of post-translational trafficking of KCNE2, a plasma membrane (PM) protein, including the early modification within the Golgi complex (a) and final localization at the PM (b) or in lysosomes (lyso) (c), without retention in the ER. B: Differential localization patterns of MCM3AP protein in the cytoplasm and nucleus. C+D: Translocation of BACH1 and CHAF1B proteins during cell cycle. C: BACH1 protein was distributed in the cytoplasm during interphase. During mitosis, most of the protein accumulated in the nucleus and wrapped around condensed chromosomes. D: CHAF1B protein localized in the nucleus during interphase, whereas after mitotic cell division, it was found in the cytoplasm of two daughter cells. Scale bars represent 10 µm. Figure modified from Figure 3 (Hu *et al.* 2006).

Cell cycle-dependent protein translocation

Many proteins are involved in regulation and progression of the cell cycle. Among the Hsa21 proteins studied here, cell cycle dependent localization patterns – correlated

with different mitotic phases indicated by chromosomal DAPI staining – were detected for MCM3AP, BACH1 and CHAF1B.

For example, the variable distribution of the minichromosome maintenance protein 3-associated protein MCM3AP in either nucleus or cytoplasm (Figure 3-9B) correlates with its role in nuclear import and acetylation of MCM3 protein. The latter is known to be an essential factor that allows the genomic DNA to undergo one single round of replication per cell cycle. The requirement for nuclear import and acetylation of MCM3 changes according to the phase of the cell cycle, which is reflected by variable distribution of MCM3AP between nucleus and cytoplasm in different cells observed on the cell array.

Regarding the transcriptional repressor protein BACH1, a complex cell division dependent localization pattern was observed. At interphase, BACH1 was distributed throughout the entire cytoplasm. During the early phase of mitosis, however, the protein accumulated in the vicinity of condensing chromatin (Figure 3-9C), suggesting that the nuclear import of BACH1 protein is highly associated with cell division.

For the chromatin assembly factor subunit CHAF1B, an inverse translocation pattern was observed: CHAF1B was translocated from the nucleus into the cytoplasm during cell division (Figure 3-9D). Consistent with a previous report, we observed that during interphase until S phase, most of the CHAF1B protein localized to the nucleus, where it is required for the assembly of histone octamers onto newly-replicated DNA. After cytokinesis, CHAF1B protein was detected in the cytoplasm of the daughter cells, probably in its inactive form after dissociation from the assembled chromatin.

3.3.5 Comparison of N- and C-terminal fusion tags

In some cases, the addition of fusion tags to the N- or C-terminus of a protein can lead to alteration of its subcellular localization. To investigate whether the observed localizations of Hsa21 proteins were significantly affected by the location and/or identity of the epitope tag, expression constructs for the production of C-terminally tagged Hsa21 proteins were constructed. For this, the ORFs of 17 Hsa21 proteins were subcloned – with Kozak sequence and without stop codon – into an expression vector encoding a C-terminal c-myc epitope (Figure 3-7B). The localizations of these proteins represented all localization categories that were previously classified using

the $(His)_6$ tag. Following the same workflow, myc-tagged proteins were expressed on HEK293T cell arrays and detected with anti-myc antibody.

The localization of C-terminally tagged proteins did not substantially differ from the localization of the N-terminally tagged proteins. One exception was the plasma membrane protein KCNJ6, a potassium channel protein which, when tagged by C-terminal myc tag, was found only in intracellular vesicles (Figure 3-10B), rather than being transported to the plasma membranes, which was the case with N-terminal $(His)_6$ tag (Figure 3-10A). Apart from this one exception, the small N-terminal hexahistidine tag appeared to be a reliable and efficient fusion tag for determination of subcellular localizations.

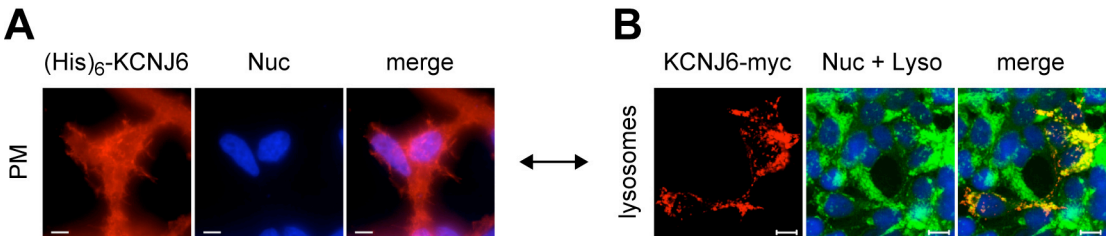

Figure 3-10. Altered localization of KCNJ6 protein resulting from different fusion tag orientation and/or identity. A: When fused with a N-terminal hexahistidine tag, the KCNJ6 protein, a subunit of the G-protein-activated potassium channel GIRK2, localized predominantly to the plasma membrane. B: When fused to a C-terminal myc tag, KCNJ6 was distributed only in intracellular vesicles, which highly overlapped with the lysosomal/endosomal compartment. Scale bars represent 10 µm. Figure modified from Figure 4.8 (Dissertation of Yuhui Hu, available at http://www.diss.fu-berlin.de/2006/577/).

3.3.6 Comparison with computational predictions

An interesting question remained whether the subcellular localizations described here for 52 Hsa21 proteins could also be predicted using available computational tools. To approach this question, two established algorithms predicting the subcellular localization of eukaryotic proteins (WoLF PSORT and ProtComp 4, available online at http://www.softberry.com/ and http://wolfpsort.org/) were supplied with the amino acid sequences of the 52 proteins with observed localizations in HEK293T cells. WoLF PSORT converts protein amino acid sequences into localization features based on sorting signals, amino acid composition and functional motifs (Horton *et al.* 2007), while ProtComp combines several methods of protein localization prediction, namely neural networks-based prediction, comparison with homologous proteins of known localization, prediction of functional peptide sequences (signal peptides, signal anchors, GPI anchors, transit peptides of

mitochondria and transmembrane segments), and search for certain localization-specific motifs (Source: www.softberry.com).

Table 3-5 lists all experimentally observed as well as computationally predicted subcellular localizations for 52 Hsa21 proteins.

Table 3-5. Comparison of observed and computationally predicted subcellular localizations of 52 Hsa21 proteins.

Gene symbol(s)	Gene function	Observed localization [a]	ProtComp 4 prediction [a]	WoLF PSORT prediction [a]
WDR9 (BRWD1_3') [b]	unknown	Nuc	Nuc	Nuc
ETS2	transcriptional regulator	Nuc	Nuc	Nuc
HMGN1	transcriptional regulator	Nuc	Nuc	Nuc
C21orf30	unknown	Nuc	Nuc	Nuc
RRP1 (NNP1)	rRNA processing	Nuc	Nuc	Nuc
WDR4	tRNA modification	Nuc	Cyto	Secr
DSCR3	unknown	Nuc	Mito	Cyto
BACH1	transcription factor	Nuc/Cyto	Nuc	Nuc
CHAF1B	chromatin assembly factor	Nuc/Cyto	Nuc	Nuc
PKNOX1	transcription factor	Nuc/Cyto	Nuc	Nuc
C21orf25	unknown	Nuc/Cyto	Nuc	Nuc
RRP1B (KIAA0179)	rRNA processing	Nuc/Cyto	Nuc	Nuc
C21orf7 (TAK1L)	unknown	Nuc/Cyto	Nuc	Cyto
PCP4	calmodulin regulator	Nuc/Cyto	Nuc	Cyto
TSGA2	chromosome-associated	Nuc/Cyto	Nuc	Secr
PCBP3	RNA binding	Nuc/Cyto	Nuc	Csk
DNMT3L	DNA methylation	Nuc/Cyto	Cyto	Nuc
MCM3AP	DNA replication inhibitor	Nuc/Cyto	Cyto	Nuc
C21orf59	unknown	Nuc/Cyto	Cyto	Cyto
PPIAL3	peptidylprolyl isomerase	Nuc/Cyto	ER	Cyto
C21orf19	unknown	Nuc/Cyto	Secr	Nuc
PDE9A	phosphodiesterase	Cyto	Nuc	Nuc
C21orf96	unknown	Cyto	Nuc	Nuc/Cyto
C21orf66 (GCFC)	DNA binding candidate	Cyto	Nuc	Cyto
CBS	cystathionine beta-synthase	Cyto	Cyto	Cyto
CCT8	chaperonin	Cyto	Cyto	Cyto
CRYZL1	unknown	Cyto	Cyto	Cyto
HSF2BP	transcriptional regulator	Cyto	Cyto	Cyto
MX1	GTP binding protein	Cyto	Cyto	Cyto
PDXK	pyridoxal kinase	Cyto	Cyto	Cyto
RPS5L	ribosomal protein S5-like	Cyto	Cyto	Cyto
SH3BGR	SH3 adapter protein	Cyto	Cyto	Cyto
UBE2G2	ubiquitination enzyme	Cyto	Cyto	Mito
HLCS	holocarboxylase synthetase	Cyto	Cyto	Nuc/Cyto
PFKL	phosphofructokinase	Cyto	Cyto	Cyto
UBASH3A	control of endocytosis	Cyto	Secr	Mito
C21orf103 (KRTAP6-1)	unknown	Cyto	Secr	Secr
TMPRSS3	transmembrane protease	ER	PM	PM
C21orf69	unknown	ER	Secr	Secr
B3GALT5	galactosyltransferase	ER/Golgi	ER	Secr
CLDN14	tight junction protein	ER/PM	PM	PM
CLDN8	tight junction protein	ER/PM	PM	PM
IFNGR2	interferon gamma receptor	ER/PM	PM	PM
AGPAT3	acyltransferase	ER/PM	PM	Secr
ABCG1	cholesterol transporter	Golgi/PM	PM	PM
CLDN17	tight junction protein	Golgi/PM	PM	PM
KCNJ15	potassium channel	Golgi/PM	PM	PM
KCNJ6	potassium channel	Golgi/PM	PM	PM
C21orf4	unknown	PM	PM	PM
CXADR	membrane receptor	PM	PM	PM
KCNE2	potassium channel	PM/Lyso	PM	PM
KCNE1	potassium channel	PM/Lyso	PM	Secr

[a] Abbreviations: Csk – cytoskeletal; Cyto – cytoplasmic; ER – endoplasmic reticulum; Golgi – Golgi apparatus; Lyso – lysosomal/endosomal; Mito – mitochondrial; Nuc – nuclear; PM – plasma membrane; Secr – secreted (extracellular).

[b] BRWD1_3' stands for a 3' fragment of the BRWD1 (WDR9) ORF, containing nucleotides 5663..6949 of RefSeq mRNA NM_033656.

As shown in Figure 3-11, a cellular localization restricted to the nucleus was correctly predicted by both algorithms for five out of seven proteins (71% correct). For evaluation of the broad category of localization to both nucleus and cytoplasm, prediction to either compartment was counted as correct in this case, resulting in 86% correct predictions by at least one program, or 71% by both programs. Regarding localization patterns restricted to the cytoplasm, WoLF PSORT performed better than ProtComp 4 (81% vs. 69% correct predictions), while for plasma membrane proteins, ProtComp 4 scored 100% correct and WoLF PSORT reached 83%. In the secretory pathway (ER and Golgi), only ProtComp 4 succeeded in predicting a localization in the endoplasmic reticulum (9% correct). None of the two programs correctly predicted a lysosomal localization. For all correct predictions, the accordance between both algorithms ranged from 83% for proteins in both nucleus and cytoplasm up to 100% for proteins restricted to the nucleus.

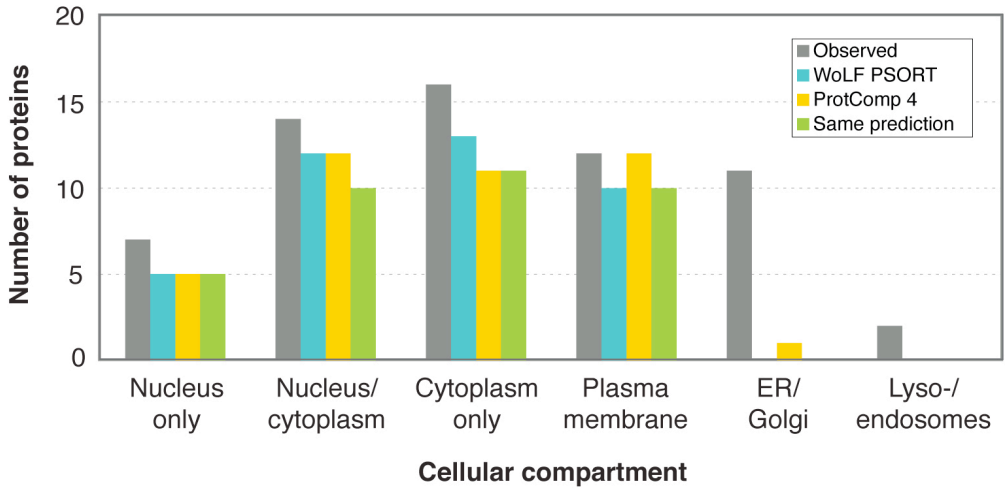

Figure 3-11. Comparison between experimentally observed and computationally predicted subcellular localizations for 52 detected Hsa21 proteins. Two different algorithms predicting the subcellular localization of eukaryotic proteins (WoLF PSORT and ProtComp 4) were feeded with a multiple FASTA file containing the amino acid sequences of the 52 proteins with observed localizations in HEK293T cells. The localizations categories from experimental observations and computational predictions were visualized in a bar diagram. ProtComp is available online at http://www.softberry.com/ and WoLF PSORT at http://wolfpsort.org/.