3.4 Identification of new Hsa21 protein-protein interactions

Building an integrated view of the macromolecular interactions involving the gene products encoded by Hsa21 represents an essential step for progressing in the understanding of Down syndrome pathogenesis. Currently, this knowledge is largely incomplete, and available information is found scattered in different public repositories, a significant part of it being referenced only in individual literature records. As a rough estimate, current large-scale human protein-protein interaction (PPI) data sets are only covering ca. 10% of the estimated human interactome (Hart *et al.* 2006). The small overlap which can be observed between the two major human Y2H screens, but also between these two sets and the PPIs recorded in literature and in PPI databases illustrate the fact that, as far as human PPIs are concerned, we are still far from reaching saturation. Moreover, the proteins encoded by Hsa21 are poorly represented in the Y2H screens reported by Rual et al. 2006 and Stelzl et al. 2006, in which only 17 and 12 Hsa21 proteins were reported, respectively (the overlap was only one protein, i.e. phosphofructokinase from liver, PFKL). Currently each novel Y2H or MS-based screen for human PPIs identifies a novel set of interactions, underlying the need for multiple and repeated screens (Causier 2004).

To expand the knowledge about protein-protein interactions of Hsa21 proteins, a set of available full ORFs was subcloned into a Y2H bait vector and used for identification of interactors in a mating-array based Y2H screen. This screen was performed in collaboration with the group of Prof. Erich Wanker at the Max Delbrück Center for Molecular Medicine (MDC) in Berlin-Buch. Interacting protein pairs identified in this screen were subjected to further validation experiments, including Y2H assays after cotransformation into fresh yeast cells, subcellular colocalization and pull-down assays in cultured mammalian cells, as described below.

3.4.1 Yeast two-hybrid constructs for interaction screening

This work was based on the set of 206 Hsa21 complete ORFs, excluding 26 paralogous KAP cluster genes. At the time when the Y2H screening procedure was initiated, 95 full ORFs were available as Gateway entry clones from the ongoing effort to clone all ORFs on Hsa21 (see above; meanwhile, 167 ORFs have been cloned, allowing for further identification of new interactions).

From the 95 available clones, 67 were chosen for Y2H screening because they contained no transmembrane domains and were not already included in the Y2H array. This set was used for recombinatorial subcloning into the Y2H bait vector pBTM116-D9, producing 62 insert-confirmed baits in fusion with the LexA DNA-binding domain (Figure 3-12A).

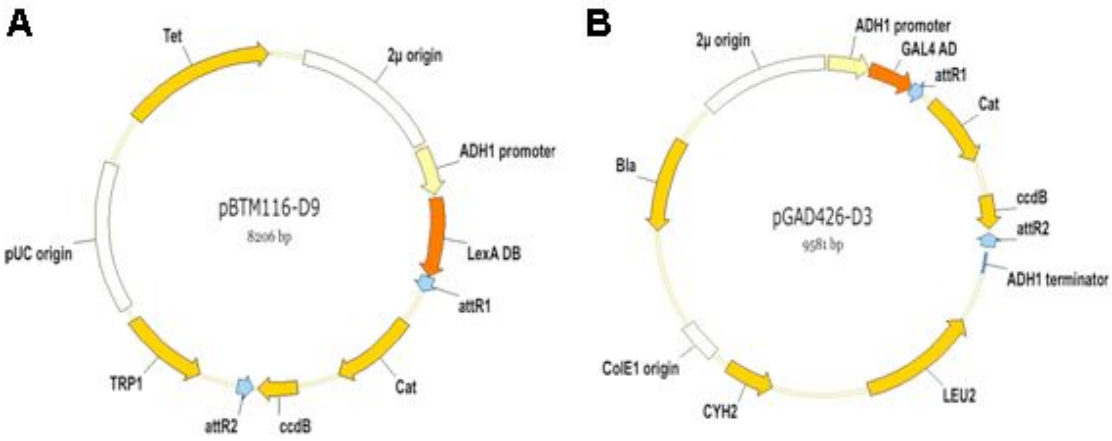

Figure 3-12. Vector maps of bait and prey vectors used for Y2H interaction assays. A: Bait ORFs were subcloned into pBTM116-D9 for yeast expression with N-terminal fusion of the LexA DNA-binding domain (LexA DB). The TRP1 growth reporter gene enabled growth of yeast transformants on minimal medium lacking tryptophan. B: Prey ORFs were subcloned into pGAD426-D3 for yeast expression with N-terminal fusion of the GAL4 transcriptional activation domain (GAL4 AD). The LEU2 reporter gene enabled growth of yeast transformants on minimal medium lacking leucine. Abbreviations: 2µ origin – yeast origin of replication; ADH1 promoter – yeast alcohol dehydrogenase promoter for expression of the recombinant fusion proteins; 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; CYH2 – cycloheximide sensitizing gene, can be used to select for only the bait ORF containing plasmid after Y2H screening; pUC/ColE1 origin – bacterial origin of replication; Bla – β-lactamase, used for Ampicillin selection and propagation of ORF-containing constructs in *E. coli*; Tet – Tetracycline resistance gene, used for selection and propagation of ORF-containing constructs in *E. coli*.

For detection of autoactivating baits, the constructs were introduced together with the empty prey vector pGAD426-D3 into yeast strain L40ccua, which carries two interaction reporters, the growth reporter URA3 and the colorimetric reporter lacZ. Additional information on this yeast strain can be found in the Appendix, chapter 6.3.2 (‘*Saccharomyces cerevisiae* strains’). Autoactivation was determined as growth on minimal medium lacking tryptophan, leucine, histidine and uracil and by colorimetric LacZ assay (see Methods). Nine bait ORFs, among them three transcription factors, were found to autoactivate one or both of the reporter genes and were not processed any further (Table 3-6). The remaining 53 bait constructs were transformed into yeast strain L40ccua and used for screening a Y2H mating array for new interactions of Hsa21 proteins, as described below.

Table 3-6. Selected properties of nine proteins autoactivating the Y2H reporter genes.

Autoactivating protein	Nucleic acid binding domain	Isoelectric point	Putative reason for autoactivation*
BACH1	BRLZ domain	4.95	DNA binding protein
ETS2	ETS domain	4.93	DNA binding protein
GABPA	ETS domain	4.91	DNA binding protein
RRP1B	Nop52 domain	9.77	RNA binding protein/ basic protein
MRPS6	-	9.30	basic protein
C21orf19	-	6.66	?
HLCS	-	5.48	?
SH3BGR	-	4.11	?
UBE2G2	-	4.86	?

*Possible reasons for the autoactivating properties of nine autoactivators were the presence of nucleic acid binding domains and/or high isoelectric points of the proteins, leading to overrepresentation of positively charged residues at the surface of the proteins, which might facilitate unspecific binding to DNA.

3.4.2 Mating array yeast two-hybrid (Y2H) screen

The 53 non-autoactivating baits were used in pools of five for mating array screening against a prey set of 5,640 ORFs fused to the GAL4 activation-domain in L40ccα yeast cells which carry the colorimetric interaction reporter lacZ. The experimental set-up has been described before (Stelzl *et al.* 2005); additional information on the yeast strain can be found in the Appendix, chapter 6.3.2 ('*Saccharomyces cerevisiae* strains'). No interaction was observed for 40 of the tested bait ORFs. Pools showing positive signals were deconvoluted by mating individual baits with the corresponding preys. As result, 29 protein-protein interactions were identified involving 13 different Hsa21 baits activating one or both reporters. 12 interacting pairs activated the growth reporter URA3 as well as lacZ, whereas 8 pairs activated only the growth reporter, and 9 pairs activated only lacZ.

In a previous large-scale screen using this mating array set-up (Stelzl *et al.* 2005), another 27 interactions had been identified for 13 different Hsa21 proteins used either as bait or prey. The overlap with the experiments presented here concerns only two Hsa21 baits (C21orf33 and PPIAL3), which however identified different interactors in the two independent screens, probably due to the fact that the focus here was on full ORF clones, in contrast to the previous study, where a large fraction of partial protein sequences was screened. In total, the mating array-based yeast two-hybrid set-up identified 56 interactions involving 24 different Hsa21 proteins (Table 3-7), which were further analyzed during the work presented here.

Table 3-7. List of 56 new protein-protein interactions involving 24 different Hsa21 proteins.

Biological process	**Hsa21 gene symbol*** (protein accession no.)	**Yeast two-hybrid interactors*** (protein accession no.)
DNA replication	MCM3AP (BAA25170.1)	**ENY2** (NP_064574.1), *CRMP1* (AAK55500.1_357..539), *MGC13379* (NP_057583.1), *NUDT5* (Q9UKK9)
histone acetylation	DNMT3L (AAH02560.1)	CSPG3 (NP_004377.2_1211..1321), WBP11 (NP_057396.1_504..641), *TLE2* (NP_003251.2)
transcription	NRIP1 (AAH40361.1)	**HTATIP** (NP_874368.1), *USP33* (AAL78314.1_414..942)
mRNA splicing	U2AF1 (AAH01923.1)	*SCMH1* (NP_036368.1_379..591), *SMC6L1* (NP_078900.1_749..1090)
mRNA transport	C21orf6 (NP_058636.1)	*DDX19* (NP_009173.1)
translation	RRP1 (AAH00380.1)	*WBP11* (NP_057396.1_504..641)
protein modification	MX1 (NP_002453.1)	*UBE2E1* (NP_003332.1)
	SUMO3 (AAH08420.1)	VIM (AAA61282.1_24..300)
	PPIAL3 (CAG32988.1)	**COPS6** (NP_006824.2), **DKFZP564O0523** (AAH33561.1), **HTATIP** (NP_874368.1), MRPL20 (NP_060441.2), WDR47 (AAH39254.1_492..893), *PRSS23* (NP_009104.1), *S100A8* (NP_002955.2), *SUPT5H* (NP_038704.1_250..1082)
metabolism	PDXK (AAH00123.1)	*WBP11* (NP_057396.1_504..641)
	CBR3 (AAH02812.1)	ARIH2 (NP_006312.1), *RAB35* (NP_006852.1)
	NDUFV3 (AAH21217.2)	*PHYHIP* (NP_055574.1)
	PFKL (AAH09919.1)	*COPS6* (NP_006824.2)
signal transduction	PCP4 (CAA63724.1)	**C1orf103** (CAD39125.1_209..493), **TP53** (NP_000537.2_248..393)
	PWP2 (NP_005040.1)	*TOM1L1* (AAH29396.1)
	TMPRSS3 (NP_076927.1)	*EEF1A1* (AAH14892.1_80..248), *RXRA* (NP_002948.1)
cell cycle regulation	BTG3 (BAA33788.1)	GIT1 (Q9Y2X7_249..761)
endocytosis	ITSN1 (AAD29952.1 1410..1721)	*EEF1A1* (AAH14892.1_80..248), *KIF5A* (NP_004975.1_731..1032), *MRPL20* (NP_060441.2), *SMARCC2* (NP_003066.2_721..1214), *SNX5* (NP_055241.1), *UNC119* (NP_473376.1)
structural component	TFF3 (AAA36766.1)	*C1orf103* (CAD39125.1_209..493), *FLJ10521* (NP_060595.2_1205..1279)
unknown	C21orf30 (CAB56001.2)	*WBP11* (NP_057396.1_504..641)
	C21orf33 (AAH03587.1)	*C10orf4* (NP_660289.2), *TLE2* (NP_003251.2), *WBP11* (NP_057396.1_504..641)
	C21orf74 (AAL82536.1)	EEF1A1 (AAH14892.1_80..248)
	C21orf91 (AAF44695.1)	**HIVEP1** (P15822_2556..2717), **HTATIP** (NP_874368.1), **WDR47** (AAH39254.1_492..893)
	C21orf127 (AAD38520.1)	**EEF1A1** (AAH14892.1_80..248), **HCNGP** (NP_037392.1), **PTN** (NP_002816.1), **TUBB2** (NP_001060.1), **UNC119** (NP_473376.1), *GTF3C1* (AAB67637.1_811..1857)

*For protein fragments, start and stop in the database sequence are indicated. Confirmation levels: (italics) solely detected in mating array Y2H set-up; (plain) detected in mating array and cotransformation Y2H set-up, where only one reporter was activated; (bold) detected in mating array and cotransformation Y2H set-up (see below), where both reporters were activated.

3.4.3 Interaction confirmation by cotransformation Y2H

The 56 interactions were validated by performing independent Y2H assays based on cotransformation of fresh yeast cells. The corresponding original prey plasmid constructs were isolated, and the prey ORFs were subcloned into Gateway vector pGAD426-D3 (Figure 3-12B). Additional information on the primers used for subcloning can be found in the Appendix, chapter 6.2.1 ('Primer for ORF cloning').

Sequence-verified plasmids were obtained for 41 PPI pairs. Then, double transformations of bait and prey constructs were performed in duplicates into fresh haploid L40ccua yeast cells. As positive control, the well-characterized interaction between Huntingtin and Endophilin A3 (alias SH3GL3) was chosen (Sittler *et al.* 1998; Ralser *et al.* 2005).

Figure 3-13 shows the results of the Y2H cotransformation assays for those protein pairs where growth on selective medium lacking tryptophane, leucine, histidin and uracil was observed, or lacZ tests gave a positive signal, indicating Y2H interaction. The amount of cell gowth varies among different interacting protein pairs, and not all interactions activating the growth reporters also result in a positive β-galactosidase test, and *vice versa*.

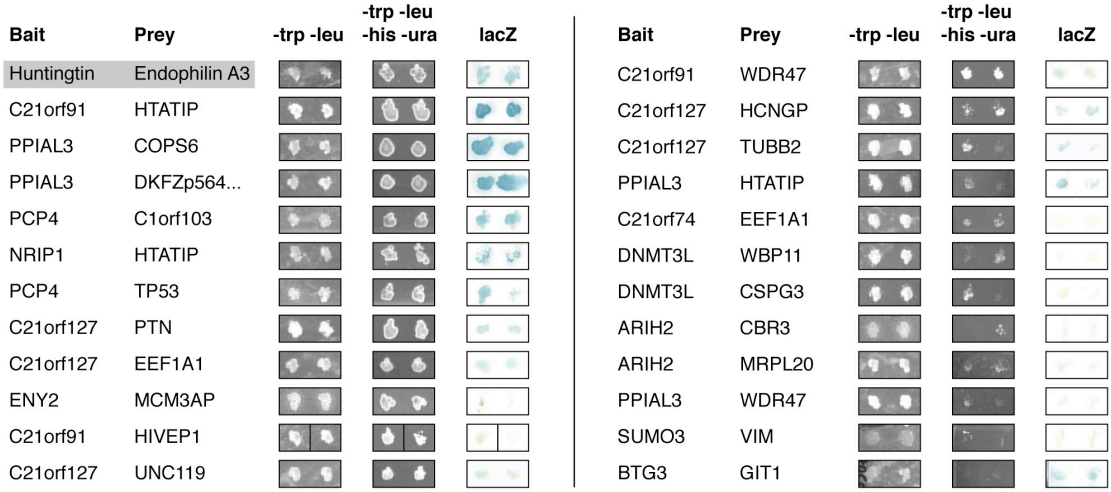

Figure 3-13. Results of a cotransformation Y2H assay for baits and preys showing positive interaction reporters in the Y2H mating array screen. Growth on medium lacking tryptophane and leucin (-trp -leu) indicates successful transformation of bait and prey vectors. Growth on medium lacking tryptophane, leucin, histidine and uracil (-trp -leu -his -ura) and blue signals in the β-galactosidase test (lacZ) indicate interaction of bait and prey. Positive control: Bait vector with Huntingtin cotransformed with prey vector containing Endophilin A3. The results are sorted according to the amount of growth observed on plates -trp -leu -his -ura. In the Y2H cotransformation asay of C21orf91 and HIVEP1, the duplicates were not replicated equally well with the 96-pin replicator; therefore, a picture from a similarly repeated experiment was added here.

Both the URA3 growth reporter and LacZ tests gave a positive result for 15 pairs of baits and preys. Only one reporter activity was observed for eight pairs: seven pairs showed growth, but a negative lacZ test, and one pair was lacZ-positive, but showed only very faint growth. Taken together, 23 interactions could be confirmed in this independent assay (56% of the 41 tested). The results of the cotransformation Y2H experiments are also present in the above Table 3-7, coded in the type style of the interaction partners: proteins in *italics* were solely detected in the mating array Y2H set-up, proteins in plain letters were detected in mating array and cotransformation Y2H set-up, where only one reporter was activated, and proteins in **bold** were detected in mating array and cotransformation Y2H set-up, where both reporters were activated.

3.4.4 Confirmation by cellular colocalization assays

The new PPIs identified by Y2H mating array screening and confirmed by cotransformation Y2H were then further analyzed using a cellular colocalization assay. A set of 16 PPIs between eight Hsa21 baits and twelve interacting preys was chosen that showed reporter activation both in mating array and cotransformation Y2H assay. Bait and prey ORFs were subcloned into mammalian expression vectors with N-terminal FLAG or HA epitope tags (Figure 3-14) and tested for expression of the recombinant proteins in COS-1 cells. Additional information on this cell line can be found in the Appendix, chapter 6.3.3 (‘*Cercopithecus aethiops* cell line’). Additional information on the primers used for subcloning can be found in the Appendix, chapter 6.2.1 (‘Primer for ORF cloning’).

Five of the eight Hsa21 baits were found expressed in the mammalian cell system, as assessed by immunodetection on Western blots. Among the constructs containing the Y2H preys, expression could be detected for six of the twelve preys. This permitted colocalization analysis for five pairs of baits and preys where both partners could be expressed in COS-1 cells (three Hsa21 proteins and four interacting proteins).

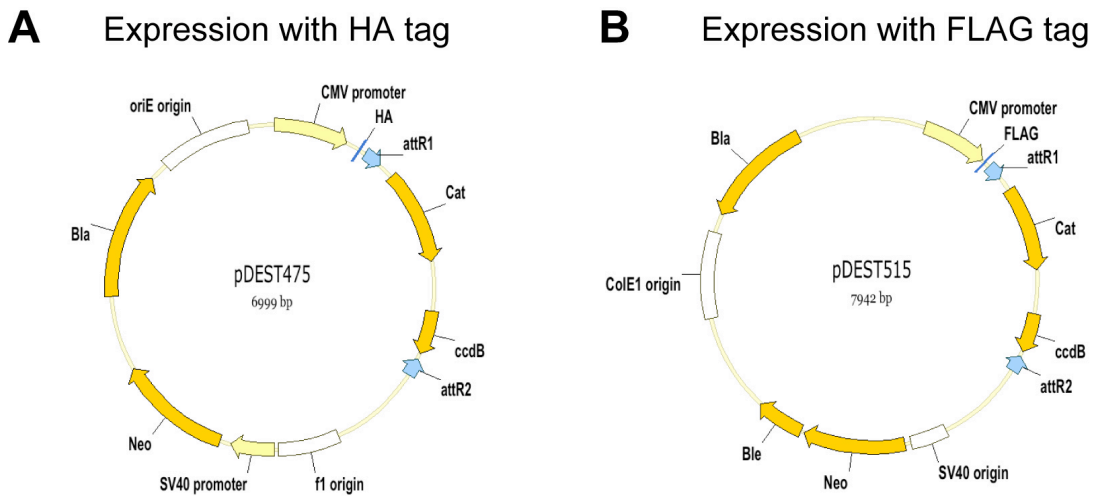

Figure 3-14. Vector maps of pDEST475 and pDEST515 used for transient fusion protein expression in COS-1 cells. A: pDEST475 was used for expression with N-terminal hemagglutinin (HA) fusion tag. B: pDEST515 was used for expression with N-terminal FLAG fusion tag. Abbreviations: CMV promoter – constitutive cytomegalovirus immediate early promoter; FLAG – N-terminal FLAG 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; f1 origin – phage f1 origin of replication, can be used for production of single-stranded DNA; SV40 promoter – constitutive simian virus 40 early promoter; SV40 origin – origin of replication derived from simian virus 40; Neo – Aminoglycoside phosphotransferase, can be used for G418 selection of stably transfected cells; Bla – β-lactamase, used for Ampicillin selection and propagation of ORF-containing constructs in *E. coli*; oriE origin – origin of replication from *E. coli*; ColE1 origin – origin of replication derived from plasmid pBR322.

Indirect immunofluorescence analysis in COS-1 cells was performed with the five coexpressed protein pairs. Among these, three protein pairs showed a colocalization in specific subcellular compartments (Figure 3-15). Peptidylprolyl-isomerase A-like 3 (PPIAL3) from Hsa21 appeared distributed at the nuclear envelope and throughout the cytosol, colocalizing with COP9 constitutive photomorphogenic homolog subunit 6 (COPS6), a component of the signalosome complex (Figure 3-15A), suggesting that PPIAL3 may be involved in cellular signal transduction via the COP9 signalosome. Hsa21-encoded Purkinje cell protein 4 (PCP4) and its Y2H interactor, the tumor suppressor protein p53 (TP53), were localized mainly to the cell nuclei, but also in the cell periphery in meshwork-like structures (Figure 3-15B). TP53 has been shown before to bind to cytoplasmic and nuclear F-actin in the presence of free calcium ions (Okorokov *et al.* 2002). PCP4, which can alter the calcium-binding properties of calmodulin by its IQ motif, might regulate the binding of TP53 to actin or its activity in a calcium-dependent manner. Hsa21 nuclear receptor interacting protein 1 (NRIP1) was mainly found in speckles at the nuclear periphery, confirming a nuclear localization as shown previously (Huq and Wei 2005). Colocalization with HIV-1 Tat interacting protein (HTATIP) was observed in cells where this speckled nuclear pattern appeared (Figure 3-15C). The two proteins, both involved in

transcriptional regulation and splicing, might act together on sites of transcription and pre-mRNA processing. Two other Y2H protein pairs tested here did not show a colocalization in COS-1 cells (PPIAL3 with WDR47 and with HTATIP).

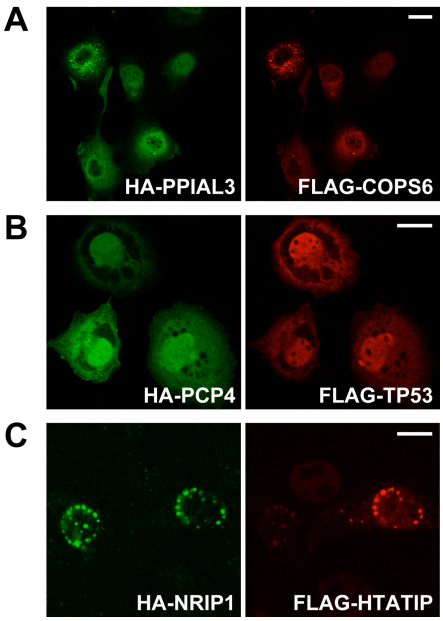

Figure 3-15. Confirmation of new interactions by colocalization assays in COS-1 cells. Immunofluorescence microscopy shows colocalizing HA- and FLAG-tagged proteins after coexpression of recombinant proteins in COS-1 cells. Color channels: (green) immunodetection with mouse monoclonal anti-HA and anti-mouse-FITC; (red) immunodetection with rabbit polyclonal anti-FLAG and anti-rabbit Cy3; (blue) nuclear DAPI staining. Scale bar = 10 µm. A: PPIAL3 and COPS6 colocalize around the nuclear envelope and throughout the cytosol; B: PCP4 and TP53 colocalize mainly in the nucleus, but also in meshwork-like structures in the cell periphery; C: NRIP1 and HTATIP colocalize in speckles at the nuclear periphery.

3.4.5 Confirmation by pull-down assays

To further confirm newly identified interactions, verification experiments were performed *in vitro* using a pull-down assay, as described before (Stelzl *et al.* 2005). Proteins with N-terminal hexahistidine fusion were produced in bacterial cells (see Methods for details). Additional information on the bacterial strain can be found in the Appendix, chapter 6.3.1 (‘*Escherichia coli* strains’). Additional information on the primers used for subcloning can be found in the Appendix, chapter 6.2.1 (‘Primer for ORF cloning’). The produced proteins were bound to Ni-NTA-beads and used for pull-down experiments with the respective interaction partners sufficiently expressed in COS-1 cells. Five protein pairs for which soluble expression of one partner could be obtained in bacteria were tested using this assay.

Three of these PPIs could be confirmed (Figure 3-16A-C), namely PPIAL3-COPS6, MCM3AP-ENY2 and PPIAL3-WDR47. For another interacting pair (PPIAL3-HTATIP), an enrichment of the binding partner was observed in respect to negative control experiments, but also some unspecific binding of the protein in the COS-1 cell lysate to the beads used in the experiment (Figure 3-16D). The fifth pair (PCP4-TP53) showed completely unspecific binding of FLAG-tagged TP53 to the beads, not allowing any interpretation (Figure 3-16E).

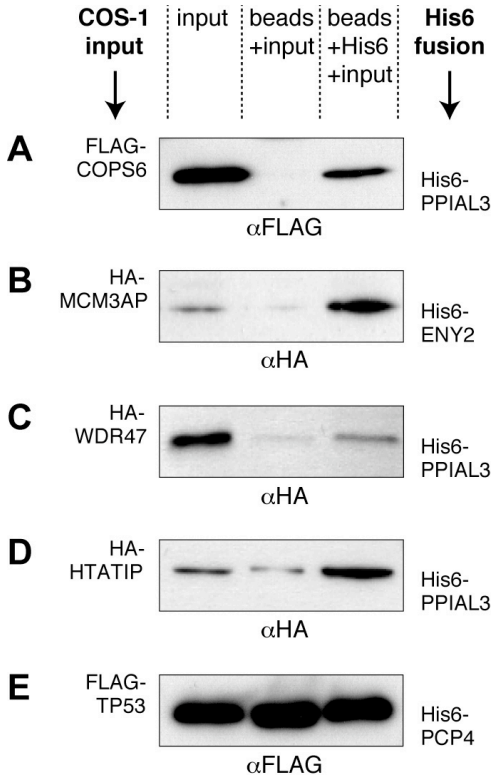

Figure 3-16. Confirmation of new interactions by pull-down assays. Pull-down experiments show *in vitro* interaction of FLAG or HA fusion proteins from COS-1 cell lysates (COS-1 input) with hexahistidine fusion proteins from bacterial cells immobilized on Ni-NTA beads (His6 fusion). Proteins were separated by SDS-PAGE, transferred to nitrocellulose membranes and detected using antibodies against the epitope indicated below each blot. Gel lanes: (input) sample of COS-1 lysate with FLAG or HA fusion protein; (beads +input) sample of control beads loaded with E. coli lysate without fusion protein and incubated with COS-1 lysate; (beads +His6 +input) sample of beads loaded with *E. coli* lysate containing His6 fusion protein and incubated with COS-1 lysate. A: COPS6 binds to PPIAL3. B: MCM3AP binds to ENY2. C: WDR47 interacts with PPIAL3. D: HTATIP binds to PPIAL3, but also weakly to blocked beads without His6 fusion protein. E: the interaction between TP53 and PCP4 cannot be analyzed using this pull-down assay because TP53 strongly binds to blocked beads without His6 fusion protein.

In summary, colocalization and pull-down experiments were performed for six different protein pairs interacting in the Y2H assays. From these, five Y2H interactions (PPIAL3-COPS6, PPIAL3-WDR47, PCP4-TP53, NRIP1-HTATIP and MCM3AP-ENY2) could be confirmed by cellular colocalization and/or pull-down assays. For an overview of the confirmation experiments and results, see Table 3-8.

Table 3-8. Overview of results obtained from PPI confirmation experiments

Hsa21 protein	Y2H interactor	Expression in COS-1	Colocalization in COS-1	Soluble expression in *E. coli*	Pull-down interaction
C21orf74	EEF1A1	- / -		- / -	
C21orf127	EEF1A1	- / -		- / -	
C21orf127	PTN	- / -		- / -	
C21orf127	UNC119	- / +		- / -	
C21orf91	HIVEP1	- / +		- / -	
C21orf91	HTATIP	- / +		- / -	
C21orf91	WDR47	- / +		- / -	
CBR3	ARIH2	+ / -		- / -	
PCP4	C1orf103	+ / -		+ / -	
PPIAL3	DKFZp564O0523	+ / -		+ / -	
MCM3AP	ENY2	+ / -		- / +	yes
NRIP1	HTATIP	+ / +	yes	- / -	
PCP4	TP53	+ / +	yes	+ / -	unspecific
PPIAL3	COPS6	+ / +	yes	+ / -	yes
PPIAL3	WDR47	+ / +	no	+ / -	yes
PPIAL3	HTATIP	+ / +	no	+ / -	yes*

Cellular colocalization could be analyzed only for protein pairs that were both expressed in COS-1 (+/+ boxes). Pull-down assays could be performed only for proteins with soluble expression in *E. coli* where the corresponding interaction partner was expressed in COS-1 cells (-/+ or +/- boxes).

*minor band from unspecific binding observed in pull-down negative control

3.5 Data collection for known protein interactions

3.5.1 Retrieval of interaction data for Hsa21 orthologous proteins

Human protein networks have been predicted based on PPIs discovered in other organisms (Lehner and Fraser 2004; Brown and Jurisica 2005; Persico *et al.* 2005). Here, to predict interactions involving Hsa21 proteins, orthology relationships among human proteins and *S. cerevisiae*, *D. melanogaster* and *C. elegans* (referred to as yeast, fly and worm) were retrieved from the InParanoid database of pairwise orthologs (Remm *et al.* 2001). By relating the orthology predictions (2,171 clusters in yeast, 5,610 in fly and 4,599 in worm) to the set of 284 genes mapping on Hsa21, 49 Hsa21 protein orthologs were identified in yeast, 105 orthologs in fly and 94 orthologs in worm (see Supplemental Table S1 available online at http://chr21.molgen.mpg.de/21ppi/S1_orthologs.xls).

The identified orthologs were used for retrieval of Y2H interaction data from large-scale Y2H screens previously performed for yeast, worm and fly (Schwikowski *et al.* 2000; Giot *et al.* 2003; Li *et al.* 2004b). The orthologous interactions are further referred to as “interologs”. Thus, 10 interologs for six Hsa21 proteins were found in yeast, 62 interologs for 25 proteins in fly and 57 interologs for seven proteins in worm. However, 45 out of the 57 worm interologs involved a ’sticky‘ protein, C03A7.4, ortholog of the keratin associated protein KRTAP10-10, and were not analyzed further. In total, 84 interologs for 31 Hsa21 proteins were obtained.

In addition, using systematic literature searches (PubMed), another 51 interologs stemming from nine different species could be identified. Taken together, 135 interologs were found for 50 different Hsa21 orthologs (Supplemental Table S2, http://chr21.molgen.mpg.de/21ppi/S2_interologs.xls). Six Hsa21 proteins were present in more than one data set, but only four overlaps in interactions (ATP5O-ATP5A1 and ATP5O-ATP5B in worm and *B. taurus*, UBE2G2-UBE2H in yeast and worm, and C21orf127-HSPC152 in yeast and worm). Comparison of all interologs with the new interactions identified in the Hsa21 Y2H screen showed one common interolog for the interaction between TMPRSS3 and EEF1A1 in fly (CG10663 with CG8280).

To identify new human interactions, the human counterparts of five Hsa21 interolog pairs, for which full ORF clones were available, were tested by co-transformation

assay in Y2H: U2AF1-SKIIP (yeast), SUMO3-UBE2D4 (yeast), PCBP3-PABPC1 (yeast), C21orf127-HSPC152 (yeast and worm), and UBE2G2-OTUB1 (fly). In this, only one interacting pair could be confirmed (C21orf127-HSPC152). Interestingly, immunofluorescence colocalization experiments showed that HSPC152 colocalized in the nucleoli of COS-1 cells with UNC119, a newly identified interactor of C21orf127 (Fig. 3-17A). This infers that C21orf127, UNC119 and HSPC152 exert a co-operating biological function in the cell nuclei.

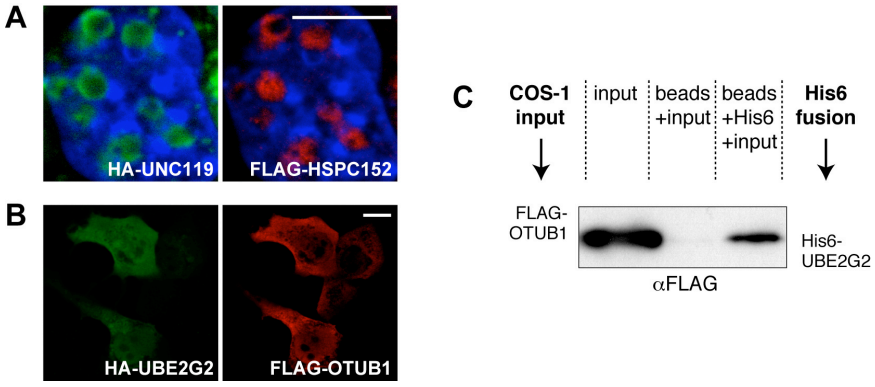

Figure 3-17. Confirmation of interologs by colocalization in COS-1 cells and pull-down assays. (A-B) Immunofluorescence microscopy shows colocalizing HA- and FLAG-tagged proteins after coexpression of recombinant proteins in COS-1 cells. Color channels: (green) immunodetection with mouse monoclonal anti-HA and anti-mouse-FITC; (red) immunodetection with rabbit polyclonal anti-FLAG and anti-rabbit Cy3; (blue) nuclear DAPI staining. Scale bar = 10 µm. A: Two new interactors of C21orf127, UNC119 and HSPC152, colocalize around DAPI-negative regions in the nucleus. B: UBE2G2 and OTUB1 colocalize in the cytoplasm. C: Pull-down experiments show *in vitro* interaction of FLAG fusion protein from COS-1 cell lysates (COS-1 input) with hexahistidine fusion protein from bacterial cells immobilized on Ni-NTA beads (His6 fusion). Proteins were separated by SDS-PAGE, transferred to nitrocellulose membranes and detected using antibodies against the FLAG epitope as indicated below the blot. Gel lanes: (input) sample of COS-1 lysate with FLAG fusion protein; (beads +input) sample of control beads loaded with *E. coli* lysate without fusion protein and incubated with COS-1 lysate; (beads +His6 +input) sample of beads loaded with *E. coli* lysate containing His6 fusion protein and incubated with COS-1 lysate. OTUB1 bound to UBE2G2 in the pull-down assay.

Another Hsa21 protein, namely UBE2G2, behaved as autoactivator in the Y2H assay. Therefore, the interaction between UBE2G2 and OTUB1 (OTU domain/ubiquitin aldehyde binding protein 1) was tested by immofluorescence experiments. Here, colocalization was observed in the cytoplasm of COS-1 cells (Fig. 3-17B), and this interaction was further confirmed with pull-down experiments (Fig. 3-17C).

Taken together, two new human interactions could be identified through interolog testing by Y2H and cellular colocalization experiments (Table 3-9). However, the interactions for the three remaining protein pairs (U2AF1-SKIIP, SUMO3-UBE2D4, and PCBP3-PABPC1) could not be verified for the human proteins.

Table 3-9. New protein-protein interactions identified through testing of interologs.

Biological process	**Hsa21 gene symbol** (protein accession no.)	**Confirmed human interaction partner*** (protein accession no.)
protein modification	UBE2G2 (AAC32312.1)	**OTUB1** (NP_060140.2)
unknown	C21orf127 (AAD38520.1)	HSPC152 (NP_057488.1)

*Confirmation levels: (plain) detected in cotransformation Y2H set-up, where both reporters were activated; (bold) further confirmed by cellular colocalization in COS-1 cells and pull-down experiments.

3.5.2 Retrieval of previously known Hsa21 interactions

A comprehensive list of all known Hsa21 PPIs was generated by combining automatic and manual data retrieval procedures when scanning the PubMed, BIND and HPRD databases and the CCSB Y2H data set (Rual *et al.* 2005). As general strategy, to circumvent the problem of heterogeneous gene nomenclature, all transcript and protein identifiers were converted to Entrez GeneIDs (see Methods for details).

Retrieval of interaction data from literature-curated databases

All human direct pair-wise PPIs were extracted from the HPRD and BIND repositories. Interactions were not taken into account if they involved (1) at least one non-human partner, (2) non-proteins partners (e.g. DNA, RNA and small molecule compounds) and (3) data referring to protein complexes. Thus, 7,705 PPIs among 4,440 different proteins were retrieved as the ‘HPRD set‘ and 4,699 PPIs connecting 2,792 different proteins as the ’BIND set‘. Using these sets, a total of 348 interactions for 68 Hsa21 proteins were extracted from HPRD and 119 interactions for 41 different Hsa21 proteins from BIND. The overlap between the two data sets was only 8% (36 out of 431 distinct PPIs derived in total). For a graphical representation of the PPI numbers and overlaps, see the Venn diagram in Figure 3-18.

Retrieval of interaction data from large-scale Y2H screens

The CCSB Y2H study (Rual *et al.* 2005) reported 30 interactions for 17 Hsa21 proteins. Of those, only two PPIs were found in HPRD and four PPIs in BIND (Figure 3-18). Six of these Hsa21 proteins were also tested in the Y2H screens reported in the work described here, where they identified other interactors. Similar observations were reported when comparing a number of genome-wide Y2H screens (Uetz and

Finley 2005). This feature may be more pronounced for the Hsa21 proteins because of the small sampling size.

Retrieval of interaction data from PubMed

Individual literature records were scanned for Hsa21 PPIs by querying the PubMed database (http://www.ncbi.nlm.nih.gov) for all annotated Hsa21 proteins (see Methods for details). This procedure resulted in a list of 303 interactions for 71 different Hsa21 proteins, the 'PubMed set'. Of those, only 32 PPIs were reported in BIND and 125 PPIs in HPRD, and 168 PPIs were not reported in any of the other interaction data sets (Figure 3-18).

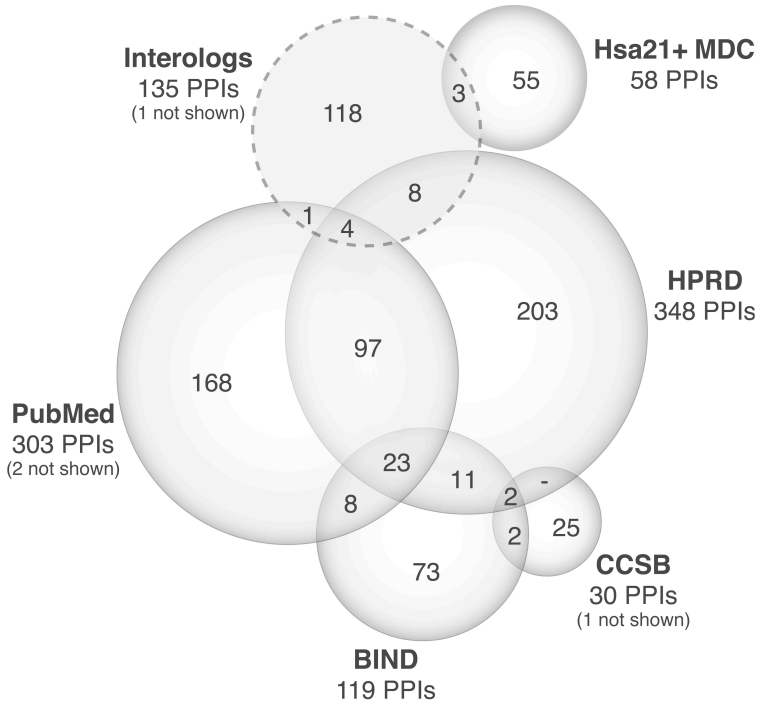

Figure 3-18. Sources and overlaps for the assembled data set of 684 PPIs of 108 Hsa21 proteins and 135 interologs from other organisms. Numbers of PPIs from each source are indicated below the source names. Circle areas are proportional to these numbers. For PPIs retrieved from more than one source, the overlap is indicated in the corresponding circle intersection. Interaction data was obtained from the following sources: (Hsa21+MDC) 58 PPIs for Hsa21 proteins, composed of 29 new PPIs from our Hsa21 Y2H screen plus 27 PPIs from the MDC Y2H screen (Stelzl *et al.* 2005) plus two new PPIs identified from interologs; (Interologs) 135 PPIs for Hsa21 orthologous proteins in other organisms; (PubMed) PPI data obtained by literature searches using the PubMed server of the U.S. National Library of Medicine; (HPRD) PPI data from the Human Protein Reference Database; (BIND) PPI data from the Biomolecular Interaction Network Database; (CCSB) PPI data from the Center for Cancer Systems Biology Y2H screen (Rual *et al.* 2005). Due to spatial restrictions in graphic representation, one overlap (Interologs+CCSB) is not shown in this Venn diagram, and one overlap between five sources (HPRD+CCSB+BIND+PubMed+Interologs) is reduced to the overlap HPRD+BIND+CCSB.

3.5.3 Consolidation and overlap of interaction data sets

Interaction data from all different data sets were entered into one comprehensive collection of all new and previously reported PPs for human chromosome 21 proteins. Combining all the data together, a set of 684 binary PPIs involving 108 different Hsa21 proteins was assembled, of which 58 new PPIs (8.5%) have been identified in the Hsa21 Y2H and interolog screens described above ('Hsa21+MDC'). As illustrated in Figure 3-18, there is relatively little overlap between the interactors identified in the different data sets.

Interestingly, there was no overlap between the results from the Hsa21 Y2H screen and previously published interactions (PubMed, HPRD, BIND). The question arose what reason there could be for these apparently false-negative interactions. A closer look into the literature showed that in almost all cases, an explanation could be found why the interaction was not re-identified in the Hsa21 Y2H screen. In literature records, 31 PPIs involving 14 Hsa21 proteins were found that had been previously shown using a Y2H assay. For eight of these 14 proteins, the reported interacting partner was not part of the prey set of the Hsa21 Y2H screen described above. Two more Hsa21 proteins are transcription factors (GABPA and ETS2) which autoactivated the reporter genes in the Hsa21 Y2H screen and consequently were not screened. Theoretically, the remaining four PPIs could have been detected in the Hsa21 Y2H screen, since the partners were present in the bait and prey sets: HSF2BP-HSF2 (Yoshima *et al.* 1998), MCM3AP-MCM3 (Takei and Tsujimoto 1998), U2AF1-NXF1 (Zolotukhin *et al.* 2002) and NRIP1-NR5A1 (Mellgren *et al.* 2003).

However, in the cases of HSF2BP, MCM3 and U2AF1, only peptide fragments had been previously tested, whereas full ORFs were used in the Hsa21 Y2H screen, a fact that may explain the differences in findings. One interacting pair (NRIP1-NR5A1) was truly missed in the Hsa21 Y2H screen, since full-length proteins were tested in both the Hsa21 Y2H screen and by Mellgren and co-workers. This conflicting result may be due to differences in the experimental set-up (e.g. different vectors and yeast strains).

A list of all PPI data collected for Hsa21 proteins during this study can be found in Supplemental Table S3 (http://chr21.molgen.mpg.de/21ppi/S3_interactions.xls). Also, the interactions can be browsed online via the dedicated '21ppi' database (http://chr21.molgen.mpg.de/21ppi), see also the next chapter.