

Systematic cloning and functional analysis  
of the proteins encoded on  
human chromosome 21

## DISSERTATION

zur Erlangung des akademischen Grades des  
*Doctor rerum naturalium*  
(Dr. rer. nat.)

eingereicht im Fachbereich Biologie, Chemie, Pharmazie  
der Freien Universität Berlin



vorgelegt von

Hans-Jörg Warnatz

aus Darmstadt

Oktober 2008

*Wissen nennen wir den kleinen Teil der Unwissenheit, den wir geordnet haben.*

zugeschrieben Ambrose Bierce (1842-1914)

1. Gutachter: Prof. Dr. Hans Lehrach
2. Gutachter: Prof. Dr. Gerd Multhaup

Disputation am 18. Februar 2009

# SHORT INDEX

<b>Summary</b> .....	I
<b>Zusammenfassung</b> .....	II
<b>Index</b> .....	IV
<b>1 INTRODUCTION</b> .....	1
<b>2 MATERIAL AND METHODS</b> .....	48
<b>3 RESULTS</b> .....	74
<b>4 DISCUSSION</b> .....	135
<b>5 BIBLIOGRAPHY</b> .....	165
<b>6 APPENDIX</b> .....	180
<b>Publikationen</b> .....	192
<b><i>Curriculum vitae</i></b> .....	193
<b>Erklärung</b> .....	195
<b>Danksagungen</b> .....	196

---

# INDEX

<b>1</b>	<b>INTRODUCTION</b>	
<b>1.1</b>	<b>Genetics, genomics and functional genomics</b>	<b>1</b>
1.1.1	The origin of genetics	1
1.1.2	The advent of genomics	2
1.1.3	The field of functional genomics	3
1.1.4	Limitations and applications of functional genomics	5
<b>1.2</b>	<b>Human chromosome 21 (Hsa21)</b>	<b>6</b>
1.2.1	Characteristics of Hsa21	6
1.2.2	Gene content and gene functions	6
1.2.3	Diseases associated to Hsa21	7
<b>1.3</b>	<b>Trisomy 21 and Down syndrome (DS)</b>	<b>11</b>
1.3.1	Incidence and impact	11
1.3.2	History and genetics	11
1.3.3	Clinical characteristics	13
1.3.4	Prenatal diagnosis and screening	21
<b>1.4</b>	<b>Molecular analysis of Down syndrome</b>	<b>23</b>
1.4.1	Defining a Down syndrome critical region (DSCR)	23
1.4.2	Mouse models for trisomy 21	24
1.4.3	Dysregulation of transcript levels	26
1.4.4	Protein dosage imbalances	28
1.4.5	Disturbed signal transduction in Down syndrome	29
<b>1.5</b>	<b>Functional Genomics Analyses of Hsa21 Proteins</b>	<b>34</b>
1.5.1	Cloning of open reading frames	34
1.5.2	Subcellular protein localization	36
1.5.3	Protein-protein interactions	40
1.5.4	Protein network analysis	43
1.5.5	Connecting interaction data with signaling pathways	45
<b>1.6</b>	<b>Objectives of this study</b>	<b>47</b>

---

<b>2</b>	<b>MATERIAL AND METHODS</b>	
<b>2.1</b>	<b>Methods in molecular biology</b>	<b>48</b>
2.1.1	Culture of <i>E. coli</i> cells	48
2.1.2	Transformation of <i>E. coli</i>	49
2.1.3	Isolation of plasmid DNA from <i>E. coli</i>	49
2.1.4	Agarose gel electrophoresis	50
2.1.5	Reverse transcription	50
2.1.6	Polymerase chain reaction (PCR)	51
2.1.7	Cloning of PCR products	54
2.1.8	DNA sequencing	54
2.1.9	Subcloning into expression vectors	55
2.1.10	Quantification of nucleic acids	56
<b>2.2</b>	<b>Methods in protein biochemistry</b>	<b>57</b>
2.2.1	Denaturing polyacrylamide gel electrophoresis (SDS-PAGE)	57
2.2.2	Coomassie staining of polyacrylamide gels	57
2.2.3	Western blotting	58
2.2.4	Immunodetection of proteins	58
2.2.5	Expression of recombinant proteins in <i>E. coli</i>	59
2.2.6	Determination of protein solubility	60
2.2.7	Quantification of proteins	60
<b>2.3</b>	<b>Methods in cell biology</b>	<b>62</b>
2.3.1	Transformation of yeast cells	62
2.3.2	Autoactivation test	62
2.3.3	Yeast two-hybrid (Y2H) assays	62
2.3.4	Culture of COS-1 cells	64
2.3.5	Transient transfection of COS-1	65
2.3.6	Immunolocalization of proteins	65
2.3.7	Pull-down assay	66
<b>2.4</b>	<b>Computational methods</b>	<b>68</b>
2.4.1	ORF primer design	68
2.4.2	Conversion of accession numbers to Entrez GeneIDs	68
2.4.3	Retrieval of orthologous interaction data (interolog data)	68
2.4.4	Retrieval of known protein-protein interactions from literature-curated databases	69
2.4.5	Retrieval of known protein-protein interactions from previous large-scale Y2H screens	70
2.4.6	Retrieval of known protein-protein interactions from PubMed	70
2.4.7	Visualization of protein interaction networks	71
2.4.8	Analysis of protein network topology	72
2.4.9	Retrieval of Gene Ontology (GO) annotations	72
2.4.10	Retrieval of pathway connections through PPI data	72

---

<b>3</b>	<b>RESULTS</b>	
<b>3.1</b>	<b>Annotation of the proteins encoded on Hsa21</b>	<b>74</b>
3.1.1	Annotation of open reading frames (ORFs)	74
3.1.2	Retrieval of transcript sequences and ORF coordinates	75
3.1.3	Primary structure of Hsa21 protein sequences	76
3.1.4	Topology predictions for Hsa21 proteins	76
<b>3.2</b>	<b>ORF cloning and protein expression</b>	<b>79</b>
3.2.1	Sources of ORF primers	79
3.2.2	Amplification and cloning of Hsa21 ORFs	79
3.2.3	Subcloning of ORFs	84
3.2.4	Protein expression in bacterial cells	85
3.2.5	Protein expression in yeast cells	87
3.2.6	Protein expression in mammalian cells	87
3.2.7	Comparison of protein expression in different cellular systems	88
<b>3.3</b>	<b>Subcellular localization of Hsa21 proteins</b>	<b>89</b>
3.3.1	Cloning of mammalian expression constructs	89
3.3.2	Antibody selection for hexahistidine and cell organelles	91
3.3.3	Determination of subcellular localizations on cell arrays	93
3.3.4	Protein translocation	98
3.3.5	Comparison of N- and C-terminal fusion tags	100
3.3.6	Comparison with computational predictions	101
<b>3.4</b>	<b>Identification of new Hsa21 protein-protein interactions</b>	<b>104</b>
3.4.1	Yeast two-hybrid constructs for interaction screening	104
3.4.2	Mating array yeast two-hybrid (Y2H) screen	106
3.4.3	Interaction confirmation by cotransformation Y2H	108
3.4.4	Confirmation by cellular colocalization assays	109
3.4.5	Confirmation by pull-down assays	111
<b>3.5</b>	<b>Data collection for known protein interactions</b>	<b>114</b>
3.5.1	Retrieval of interaction data for Hsa21 orthologous proteins	114
3.5.2	Retrieval of previously known Hsa21 interactions	116
3.5.3	Consolidation and overlap of interaction data sets	118
<b>3.6</b>	<b>Assembly and analysis of Hsa21 interaction networks</b>	<b>119</b>
3.6.1	Computational generation of a Hsa21 interaction network	119
3.6.2	Functional analysis of interaction subnetworks	121
3.6.3	Prediction of new protein functions from interaction networks	123
3.6.4	Retrieval of pathways with involvement of Hsa21 proteins	132

---

<b>4</b>	<b>DISCUSSION</b>	
<b>4.1</b>	<b>On the annotation of genes and ORFs on Hsa21</b>	<b>135</b>
4.1.1	Protein-coding versus non-protein-coding genes	135
4.1.2	On the quality of the gene annotations used in this study	136
4.1.3	On the quality of the ORF annotations used in this study	136
<b>4.2</b>	<b>ORF cloning as resource for functional genomics</b>	<b>137</b>
4.2.1	Approaches in functional genomics and proteomics	137
4.2.2	Hsa21 ORF cloning for clone-based proteomics	138
<b>4.3</b>	<b>New subcellular localizations for Hsa21 proteins</b>	<b>139</b>
4.3.1	Different small fusion tags resulted in similar localizations	139
4.3.2	Success rate of subcellular localization experiments	140
4.3.3	Observed localizations and potential effects of epitope tagging	140
4.3.4	Known and new localizations in different compartments	141
4.3.5	Protein translocation	147
4.3.6	Comparison with computational predictions	148
<b>4.4</b>	<b>New protein interactions for Hsa21 proteins</b>	<b>149</b>
4.4.1	On the identification of new PPIs by mating array-based Y2H screening	149
4.4.2	On the verification of protein-protein interactions	151
<b>4.5</b>	<b>On current protein-protein interaction data sets</b>	<b>152</b>
4.5.1	Retrieval of PPI data from databases and large Y2H screens	152
4.5.2	Retrieval of PPI data from literature records	153
4.5.3	Characteristics of proteins without PPI information	153
<b>4.6</b>	<b>On the analysis of protein interaction networks</b>	<b>154</b>
4.6.1	Higher-order protein networks	154
4.6.2	Mapping protein interactions to cellular processes	155
<b>4.7</b>	<b>Interaction networks and signaling pathways</b>	<b>156</b>
4.7.1	Previously known pathway connections of Hsa21 proteins	157
4.7.2	Newly identified pathway connections of Hsa21 proteins	161
<b>4.8</b>	<b>Conclusion and outlook</b>	<b>163</b>

<b>5</b>	<b>BIBLIOGRAPHY</b> .....	<b>165</b>
<b>6</b>	<b>APPENDIX</b>	
<b>6.1</b>	<b>Sources of cloned ORFs</b> .....	<b>180</b>
6.1.1	ORFs cloned from public cDNA clones .....	180
6.1.2	ORFs cloned from cDNA .....	182
6.1.3	ORFs cloned from genomic DNA .....	183
6.1.4	ORFs obtained from public Gateway clones .....	184
<b>6.2</b>	<b>List of primer sequences</b> .....	<b>185</b>
6.2.1	Primer for ORF cloning .....	185
6.2.2	Primer for C-terminal epitope tagging .....	188
6.2.3	Primer for insert verification and sequencing .....	189
<b>6.3</b>	<b>Bacterial and fungal strains and mammalian cell lines</b> .....	<b>189</b>
6.3.1	<i>Escherichia coli</i> strains .....	189
6.3.2	<i>Saccharomyces cerevisiae</i> strains .....	190
6.3.3	<i>Cercopithecus aethiops</i> cell line .....	190
6.3.4	<i>Homo sapiens</i> cell line .....	191
	<b>Publikationen</b> .....	<b>192</b>
	<b>Curriculum vitae</b> .....	<b>193</b>
	<b>Erklärung</b> .....	<b>195</b>
	<b>Danksagungen</b> .....	<b>196</b>



---

## DANKSAGUNGEN

Bei meinen Betreuern Dr. Marie-Laure Yaspo und Prof. Dr. Hans Lehrach möchte ich mich für die interessante und ideenreiche Zusammenarbeit bedanken. Auch möchte ich Prof. Dr. Gerd Multhaup für die freundliche Bereitschaft zur Übernahme des zweiten Gutachtens danken.

Daniela Balzereit und Sabine Schrunner haben mich tatkräftig bei einigen umfangreichen Laborarbeiten unterstützt, vielen Dank für diese Hilfe.

Auch bei allen meinen Kollegen möchte ich mich für zahlreiche interessante Diskussionen, Ideen und schöne Zeiten bedanken, vor allem natürlich bei Robert Querfurth und Dr. Marc Sultan.

Mein Dank geht auch an alle Kollaborationspartner, vor allem an Dr. Bernhard Korn (ORF-Klonierung), Dr. Yuhui Hu und Dr. Michal Janitz (Transfected Cell Arrays), Dr. Ulrich Stelzl, Dr. Martin Strödicke und Prof. Dr. Erich Wanker (Y2H Screening) sowie an Reha Yildirimman, Hyeong Jun An und Dr. Michael Lappe (Bioinformatische Analysen).

Nicht zuletzt möchte ich meiner Familie für die Unterstützung danken, vor allem meinen beiden Eltern, die immer für mich da waren. Und Dank auch an Sanna für das Verständnis, wenn es mal spät wurde.