

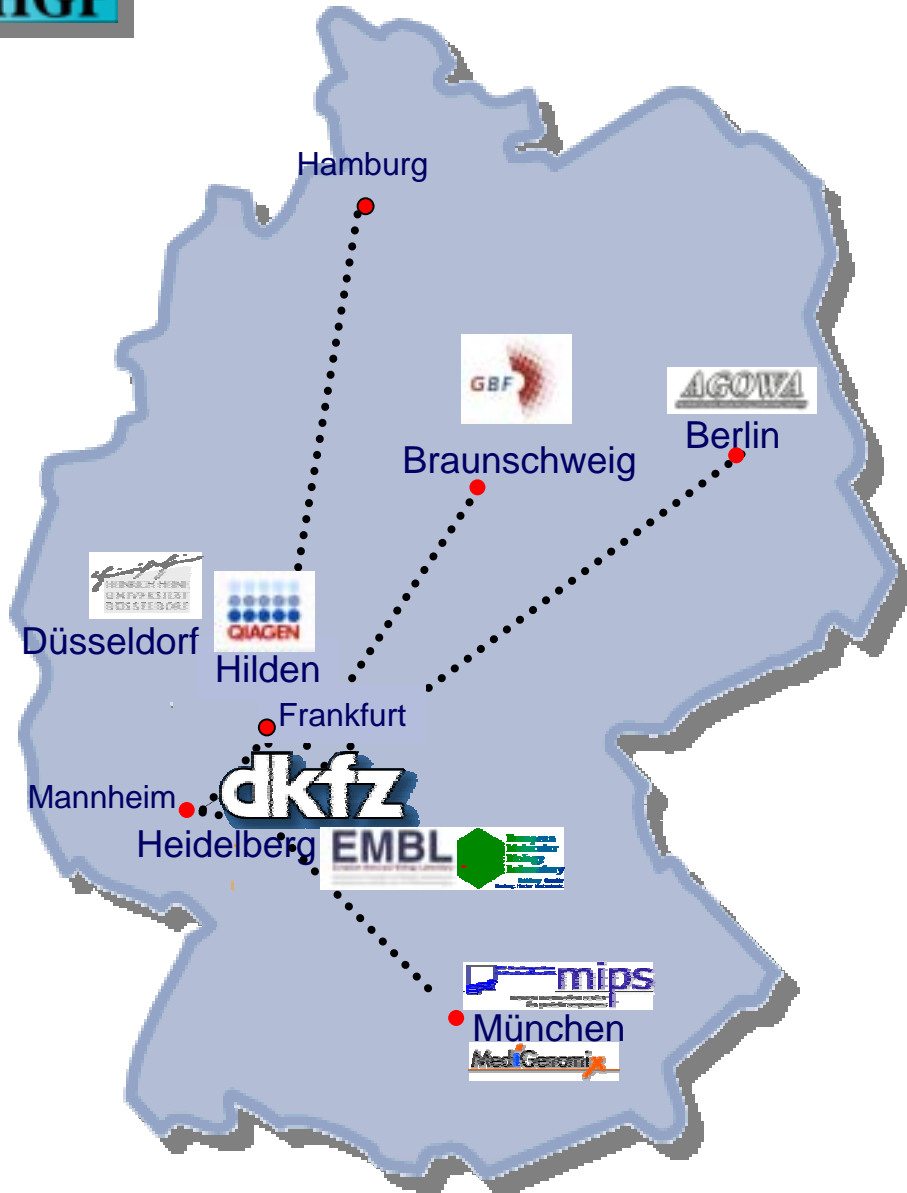
The German cDNA Network – cDNAs for Functional Genomics and Proteomics

Stefan Wiemann

Dept. of Molecular Genome Analysis
DKFZ Heidelberg, Germany



full-length cDNAs The German cDNA Network



coordination

-

cDNA Resources

-

sequencing

-
-
-
-
-
-
-

bioinformatics

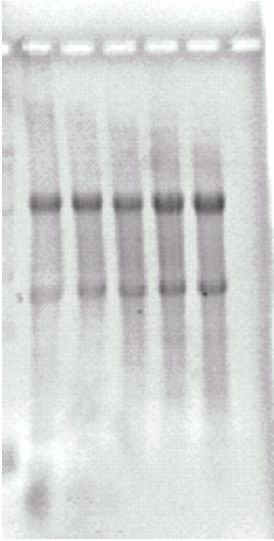
-
-

functional genomics

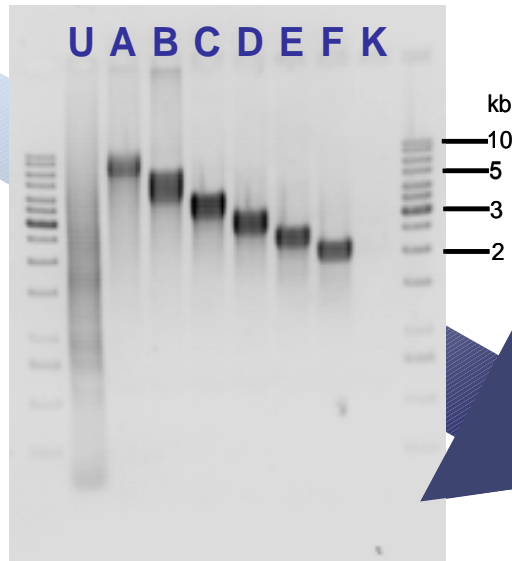
-
-

cDNA library technology at the DKFZ

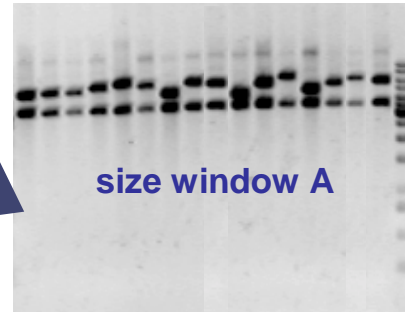
high quality tissues / RNA from collaborating pathologists



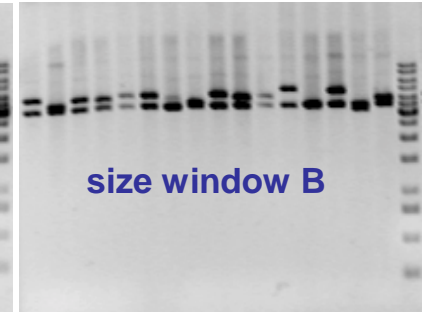
long range first strand synthesis



size fractionation of uncloned cDNA
U = unfractionated cDNA, A-F = size fractions

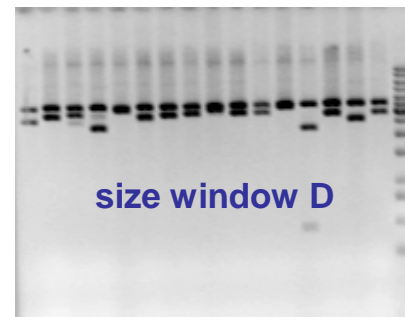


size window A

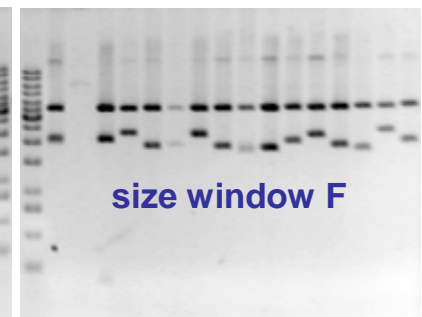


size window B

analysis of cloned size fractionated cDNA
cloning vector: 3.7 kb



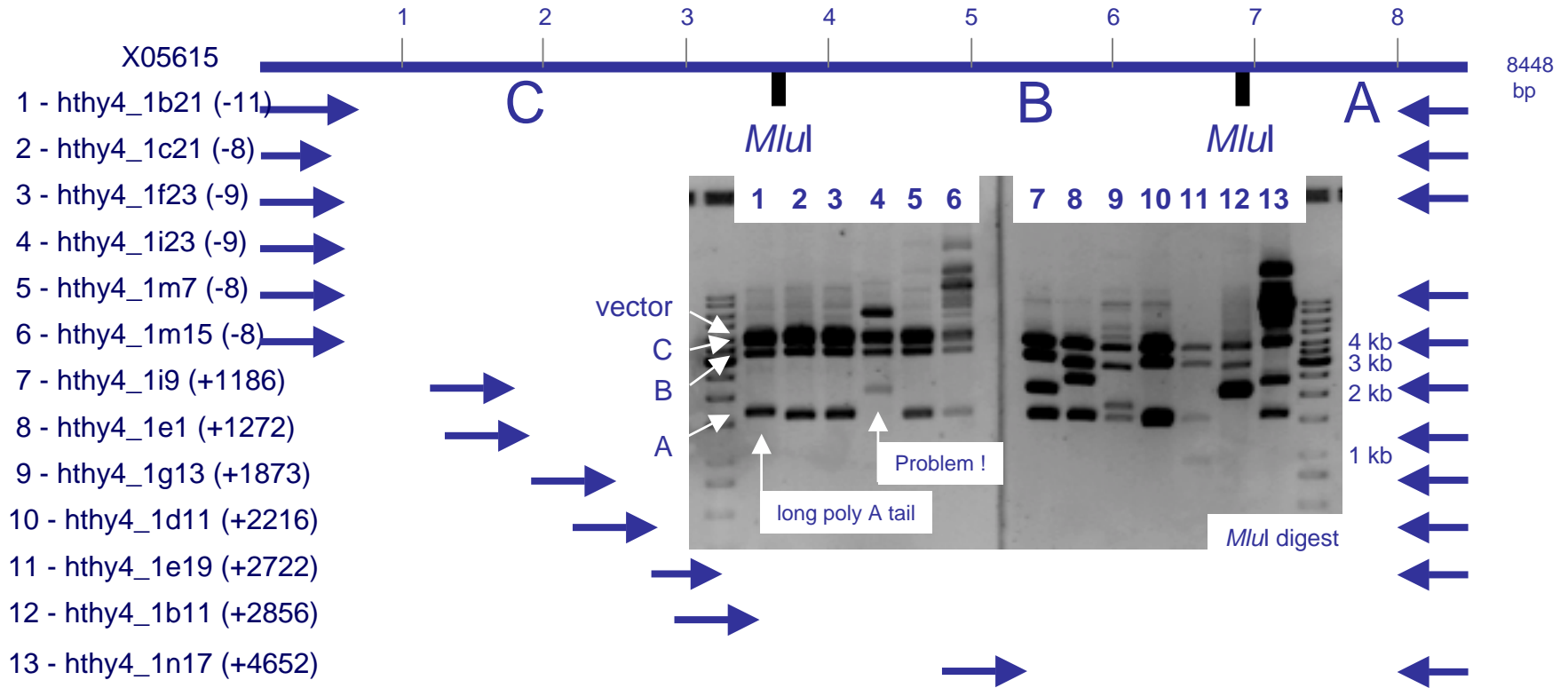
size window D



size window F

see Ruth Wellenreuter's poster

size distribution of thyroglobulin cDNAs in thyroid gland sub-library



5 / 13 clones are full-length (8.5 kb mRNA)

status (Mar 8, 2002)

		Transcriptome 2000
EST clones	100,343 clones	45,469
cDNAs	5,263	3,713
	14,461,397 bp	8,986,305
av.insert size	2,750 bp	2,420

all clones are available through clone@rzpd.de (Resource Center)
www.rzpd.de

The Genomic Organization of the Gene Encoding the Vanilloid Receptor: Evidence for Multiple Splice Variants

Qing

GENOMES 34, 256-266 (1998)
ARTICLE NO. GENES34

Genomic Organization of a 225-kb Region in Xq28 Containing the Gene for X-Linked Myotubular Myopathy (*MTM1*) and a Related Gene (*MTMR1*)

Gao

Group 2

*Dedong Chen
Patrick J. J. van
Gennip
RabierThe American Society for Human Genetics
© 1998 by The American Society for Human Genetics, Inc.

Identification and cDNA Cloning of a Novel RNA-binding Protein That Interacts with the Cyclic Nucleotide-responsive Sequence in the Type-1 Plasminogen Activator Inhibitor mRNA*

Vol. 176, No. 3, June 15, 1998, pp. 584-592, 2001
Printed in U.S.A.Jin
Ph
Li

FEBS 2771

FEBS Letters 460 (1999) 153-160

DelGEF, an RCC1-related protein encoded by a gene on chromosome 11p14 critical for two forms of hereditary deafness¹

*Olivier
*David

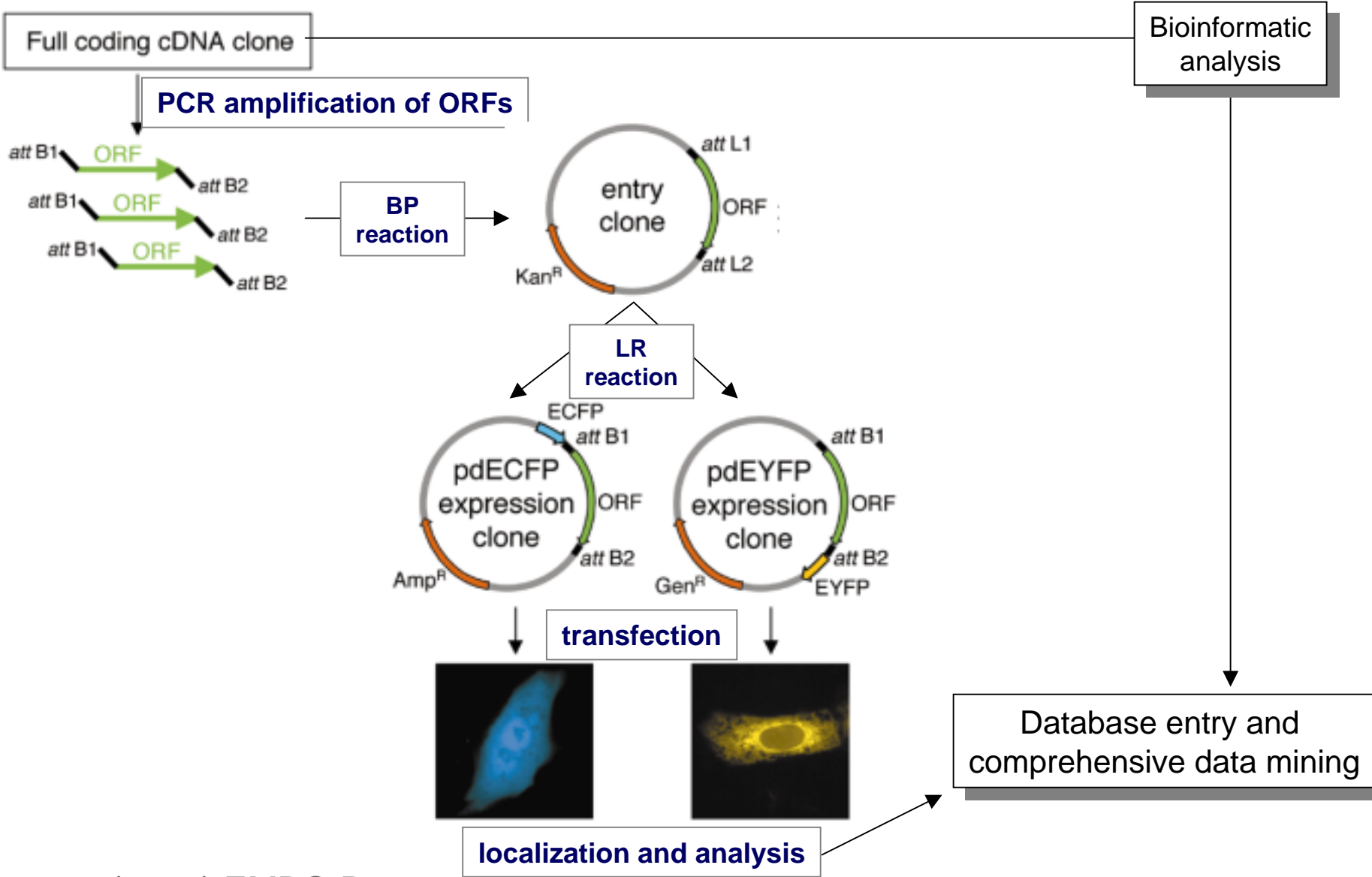
Mokobé J. DORVILLE, Patrick A. Lind

343

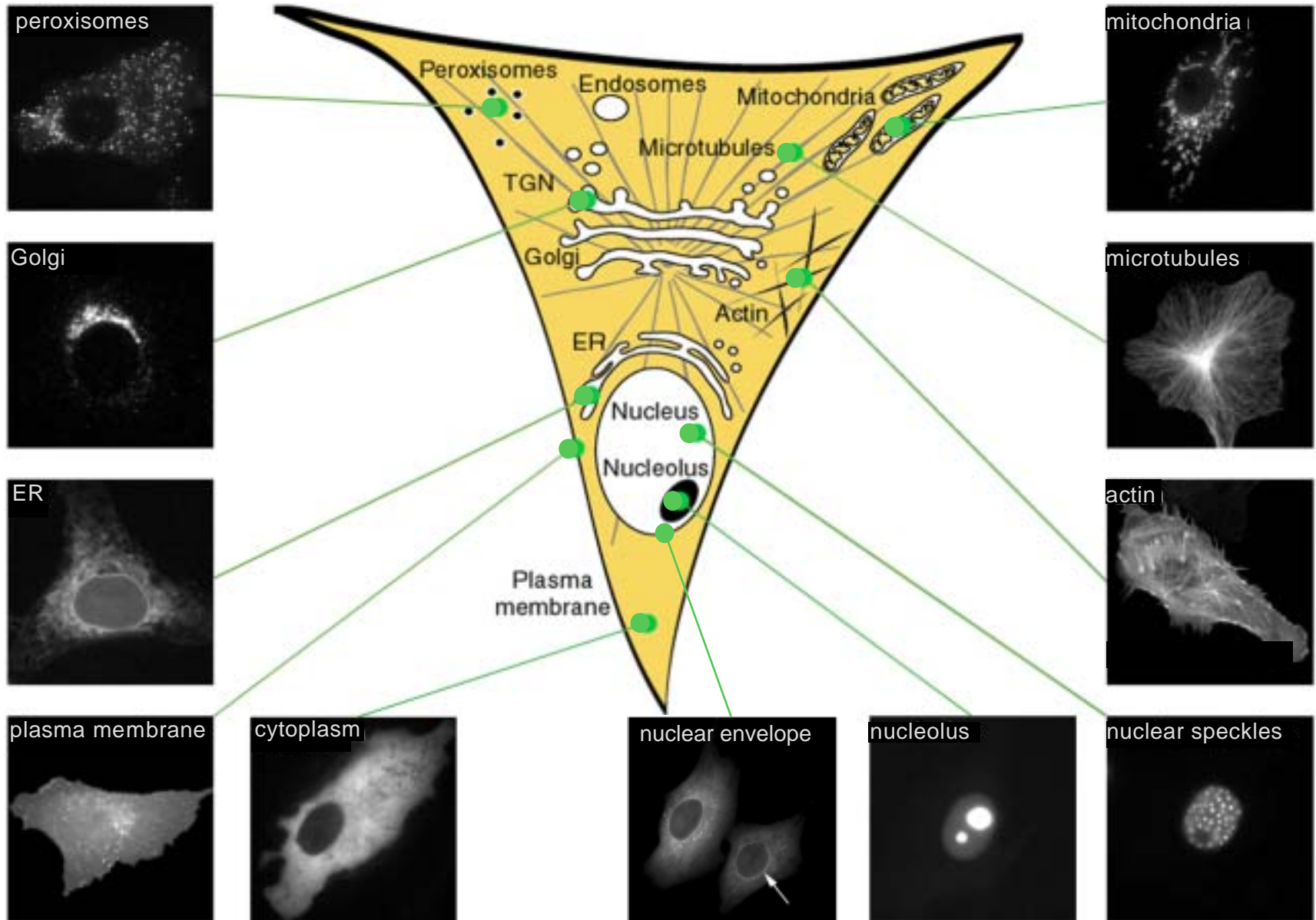
Cloning and expression of a cDNA encoding human inositol 1,4,5-trisphosphate 3-kinase C

Valérie DENWASTÉ*, Valérie ROUILLOU*, Colette MOREAU*, Stephen SHEARS†, Kazunaga TAKAZAWA‡ and Christophe ERMEJOU*
*Hôpital Raymond Poincaré (APHP), INSERM UMR de Biologie, Campus Cochin, Bldg. C, 808 Route de Lenoir, 91191 Evry-sur-Seine, France; †Hôpital
Signaling Systems, National Institute of Environmental Health Sciences, Research Triangle Park, NC 27709, U.S.A., and ‡Genentech Molecular, Tokyo Yokohama Hospital,
Tokyo 100, Japan

systematic cloning of full-length open reading frames



cDNAs – tools for protein localization



<http://www.dkfz-heidelberg.de/abt0840/GFP>



MOLECULAR GENOME ANALYSIS



Search

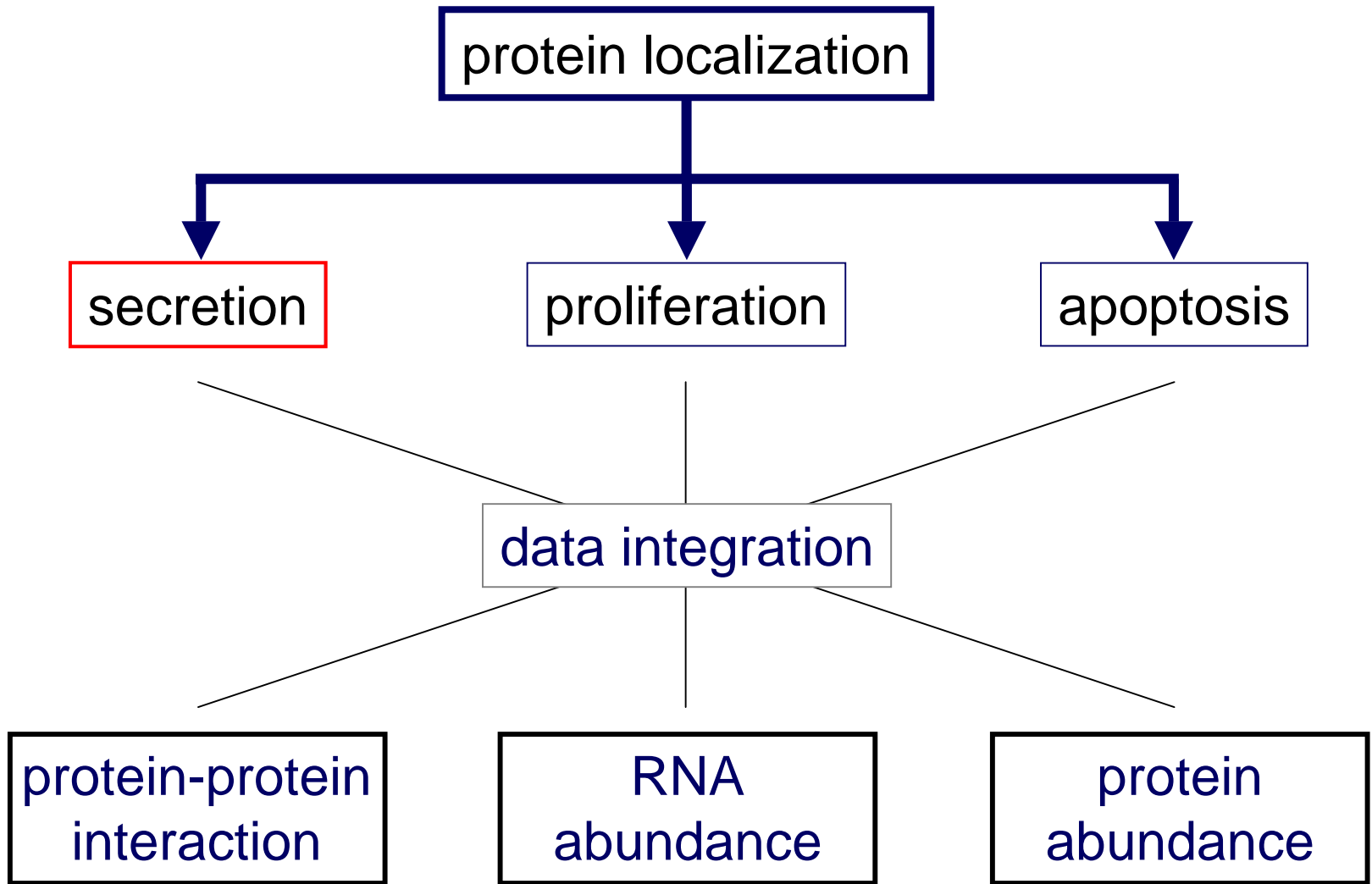
- Welcome
- Overview
- Strategy
- Statistics
- Data table
- Controls
- Publications



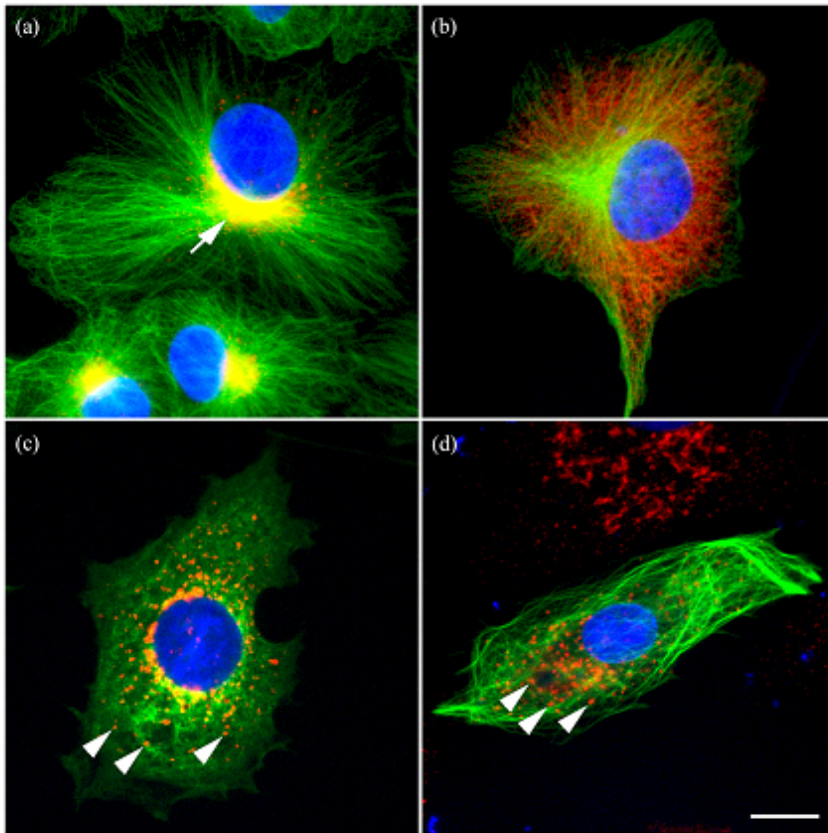
Chrom	Location	STS	Localization	Predicted	Localization	Picture			
	512.1 cR from top of Chr10 linkage group		"secre pathway"	Cytosol + Nucleus	Picture				
	333.4 cR from top of Chr16 linkage group			Cytosol	Picture				
	194.2 cR from top of Chr15 linkage group			Cytosol + Nucleus	Picture				
	2q12.1			Cytosol	Picture				
	18			Nucleus	Picture				
	10			None	Picture				
DKFZp434J0450	AL136869	1180	36	569	178	unknown	unknown	Nucleus	Picture
DKFZp434J154	AL80155	2383	218	1579	454	unknown	unknown	Cytosol	Picture
DKFZp434L134	AL136865	1960	98	1516	473	nucleic acid management	transcription factor with two C2H2 zinc fingers, similar to the mouse Zfp-38 protein that regulates transcription during spermatogenesis	Nucleus	Picture
DKFZp434N0250	AL117525	1584	154	1539	462	cell cycle	protein kinase	Cytosol + Nucleus	Picture
DKFZp434N0535	AL117518	4055	126	4025	1300	differentiation & development	similar to: Drosophila chromatin protein	None	
DKFZp434N214	AL136866	3010	456	2429	658	nucleic acid management	RNA-binding protein/RNP	None	Picture
DKFZp434O194	AL136867	3767	134	3673	1180	structure & motility	mucin	Mitochondria	Picture
DKFZp434P1735	AL136859	1966	98	1612	505	transport and traffic	BETA-CATENIN	Cytosol + Nucleus	Picture



cell biology - assays



protein action – effect of overexpression

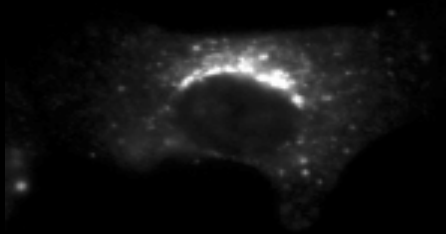


- a. **normal cell**
- b. **brefeldin A treated**
disassembles Golgi complex
- c. **nocodazole treated**
disassembles microtubules and also the Golgi complex
- d. **novel GFP-tagged ORF**
disassembles Golgi complex without affecting the microtubules

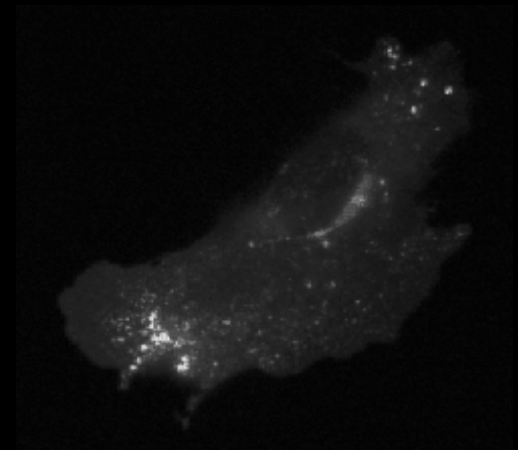
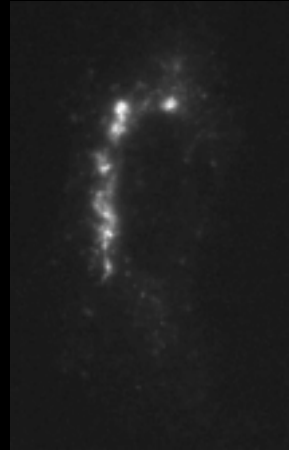
blue - nucleus (DAPI stain)
red - Golgi complex (anti COPI)
green - microtubules (anti tubulin)

protein dynamics

static view:



dynamic view:

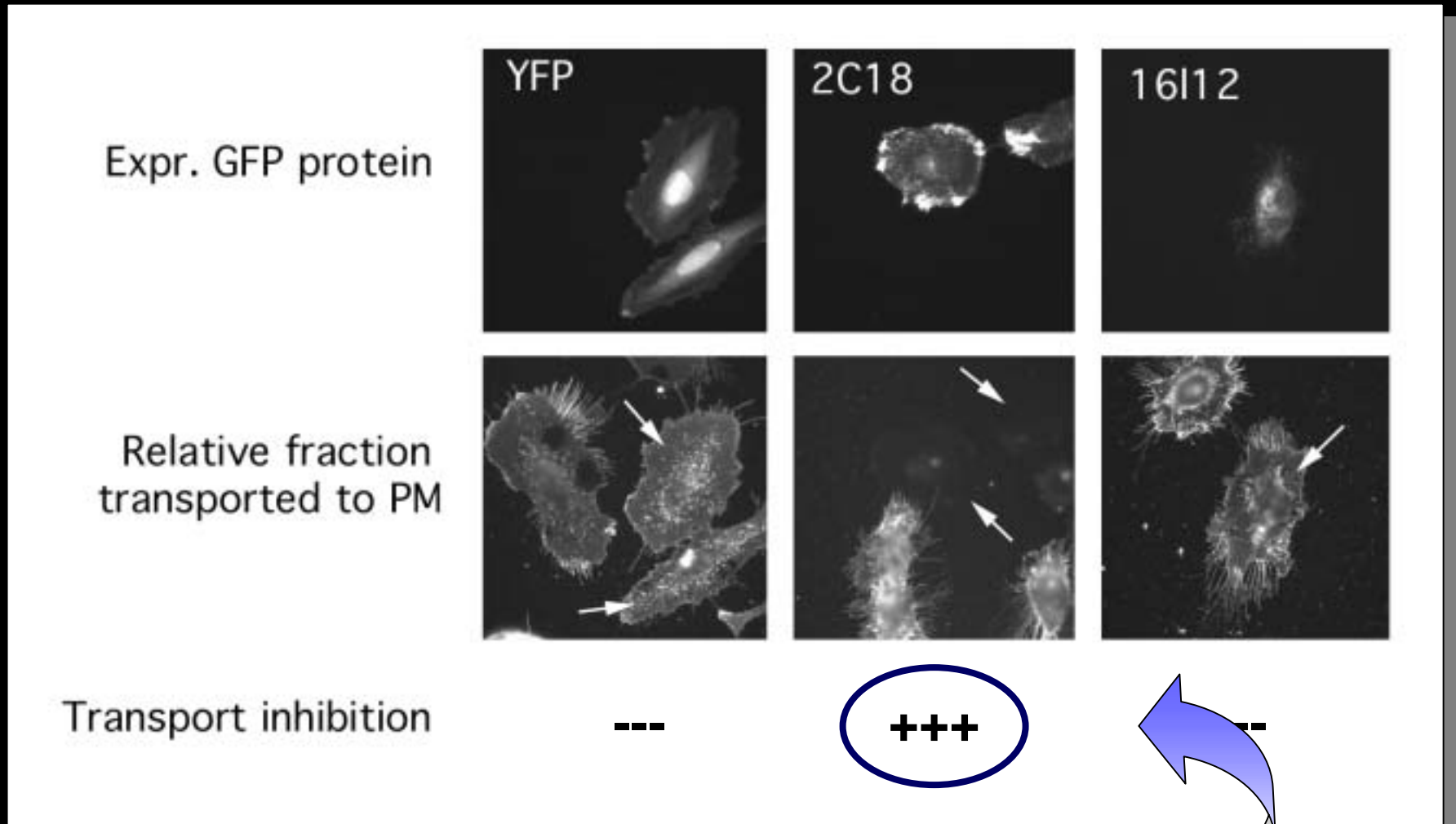


observation: protein shuttles between Golgi and plasma membrane

next question: does overexpression / repression of the protein affect transport ?

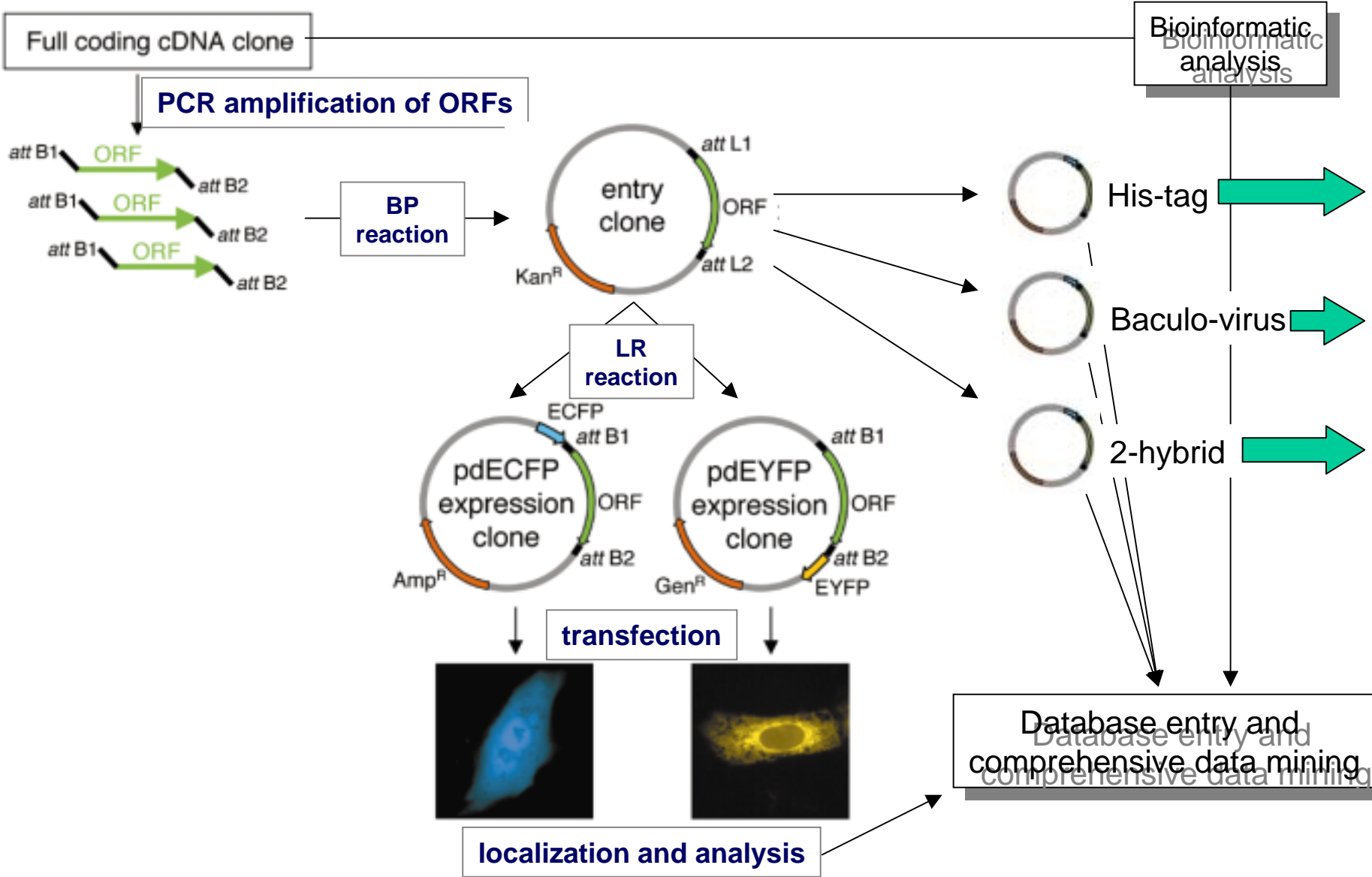
protein dynamics

conditional knock in

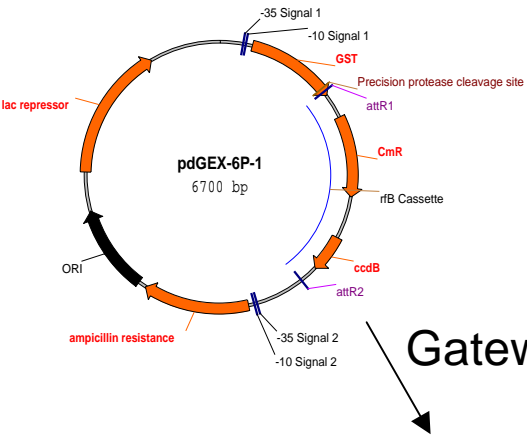


next question: does overexpression / repression of the protein affect transport ?

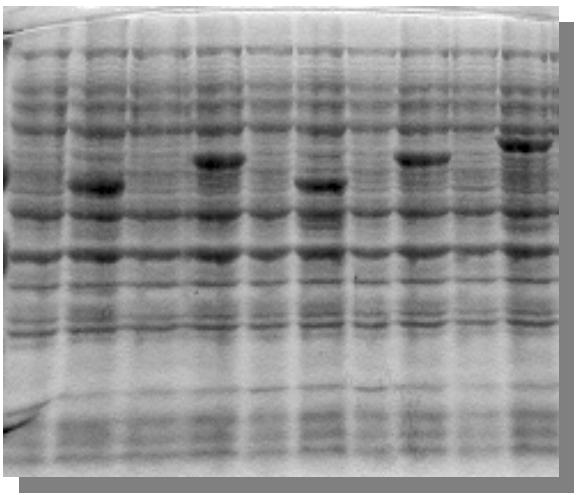
systematic cloning of full-length open reading frames



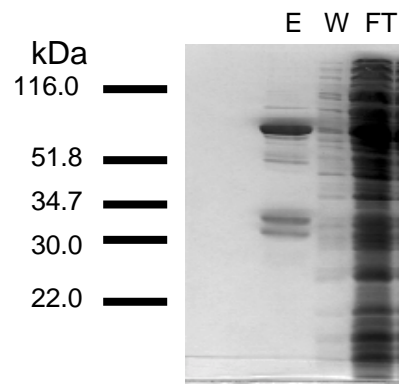
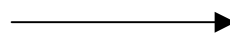
purification GST fusion protein



Gateway expression clone



GST column



- cut out protein band
- washing/shrinking
- Trypsin digest
- load sample on MALDI Target
- run MassSpec

Protein Candidates

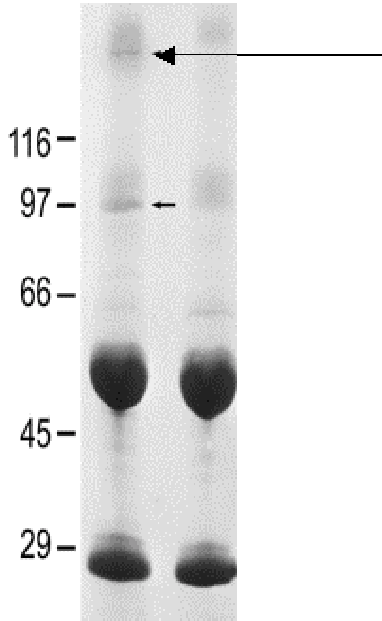
Rank	Probability	Protein Description	Species	MW (kDa)
1	4.8e-02	gi208443 gb AAB59734.1 (M97937) glutathione transferase [unidentified cloning vector] & gi12053297 emb CAB66834.1 (AL136900) hypothetical protein [Homo sapiens]	Unidentified cloning & Homo sapiens	26 & 70
2	4.8e-02	gi121697 sp P08515 GT26_SCHJA GLUTATHIONE S-TRANSFERASE 26 KDA (GST 26) (SJ26 ANTIGEN) (GST CLASS-ALPHA) & gi12053297 emb CAB66834.1 (AL136900) hypothetical protein [Homo sapiens]	Schistosoma japonicu & Homo sapiens	25 & 70
3	4.3e-02	gi3184404 dbj BAA28713.1 (AB014641) GST-stuffer fusion protein [Cloning vector pGEX-PUC-3T] & gi12053297 emb CAB66834.1 (AL136900) hypothetical protein [Homo sapiens]	Cloning vector pgex- & Homo sapiens	33 & 70
4	4.0e-02	gi1527195 gb AAB88910.1 (U67875) glutathione S-transferase [Expression vector pESP-I] & gi12053297 emb CAB66834.1 (AL136900) hypothetical protein [Homo sapiens]	Expression vector pe & Homo sapiens	28 & 70
5	3.4e-02	gi595718 gb AAA57095.1 (U13852) glutathione S-transferase [unidentified cloning vector] & gi12053297 emb CAB66834.1 (AL136900) hypothetical protein [Homo sapiens]	Unidentified cloning & Homo sapiens	27 & 70
6	3.4e-02	gi1699065 gb AAB37349.1 (U78873) glutathione S-transferase [Cloning vector pGEX-6P-2] & gi12053297 emb CAB66834.1 (AL136900) hypothetical protein [Homo sapiens]	Cloning vector pgex- & Homo sapiens	28 & 70
7	3.4e-02	gi595706 gb AAA57086.1 (U13849) glutathione S-transferase [unidentified cloning vector] & gi12053297 emb CAB66834.1 (AL136900) hypothetical protein [Homo sapiens]	Unidentified cloning & Homo sapiens	27 & 70
		gi708205 gb AAB473682.1 (M21676) glutathione transferase [unidentified cloning	Unidentified	

GST fusion

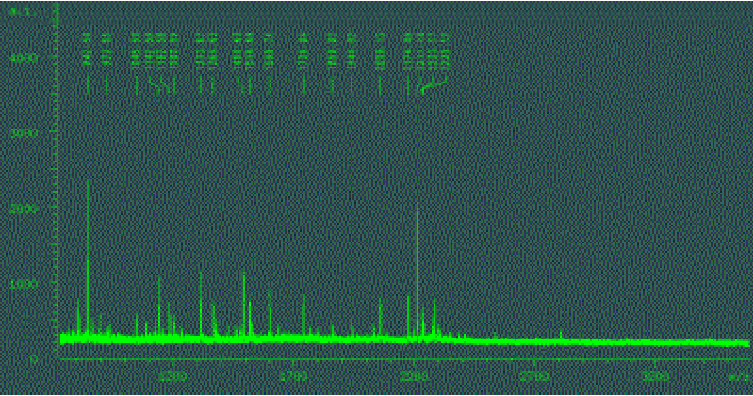
DKFZp434M108

immuno-precipitation / protein complexes

protein-protein interactions

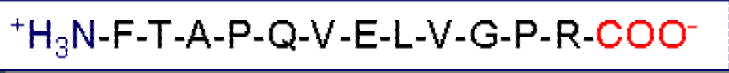
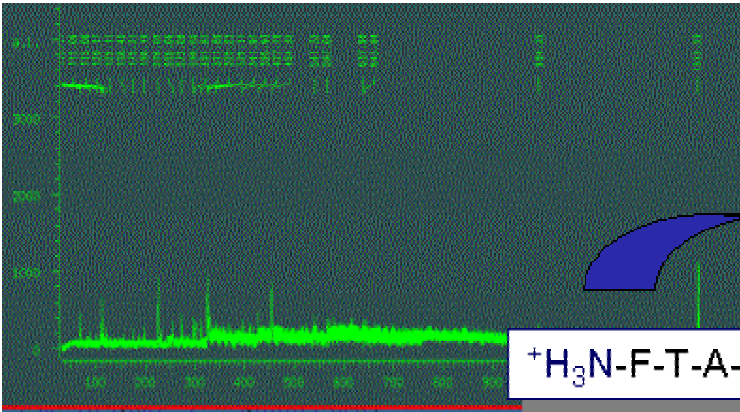


MALDI Fingerprint



Not informative

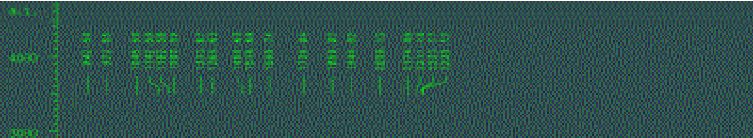
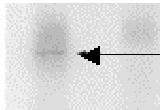
PSD spectrum



immuno-precipitation / protein complexes

protein-protein interactions

MALDI Fingerprint

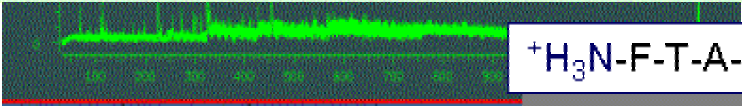


Detailed Results

Rank	# Unmatched Ions	Sequence	MH ⁺ Calculated (Da)	MH ⁺ Error (Da)	Protein MW (Da)/pI	Species	NCBI nr. 11.24.01 Accession #	MS-Digest Index #	Protein Name
1	1/16	(R) FTAPQVELVGP (L)	1313.7218	0.0218	148428.1 / 8.41	HOMO SAPIENS	10047317	546217	(AB046840) KIAA1620 protein

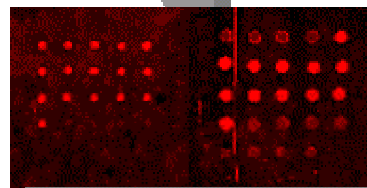
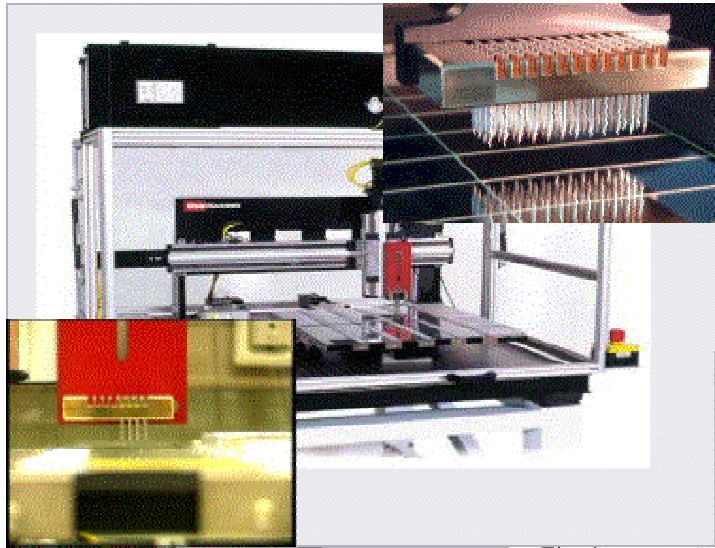
Fragment-ion (m/z)	158.46	175.21	198.10	226.02	255.00	272.08	296.86	301.81	325.22	341.97	395.92	427.78	454.32	541.32	567.08	994.25
Ion-type		y1	PQ-	PQ	y2-	y2	APQ	b3-	PQV	ELV		y4		y5		y9
Delta Da		0.09	28 -0.02	-0.10	NH ₃ -0.15	-0.09	-0.30 PQV- 28 -0.33	H ₂ O -0.34	0.03	-0.23	APQV -0.30	-0.48	PQVE 0.09	-0.03	PQVEL -0.23	-0.32

MS-Tag 4.2.1, ProteinProspector 3.2.1
 © Copyright (1995-1999) The Regents of the University of California

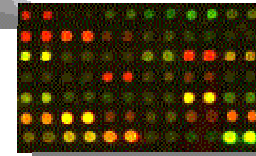


⁺H₃N-F-T-A-P-Q-V-E-L-V-G-P-R-COO⁻

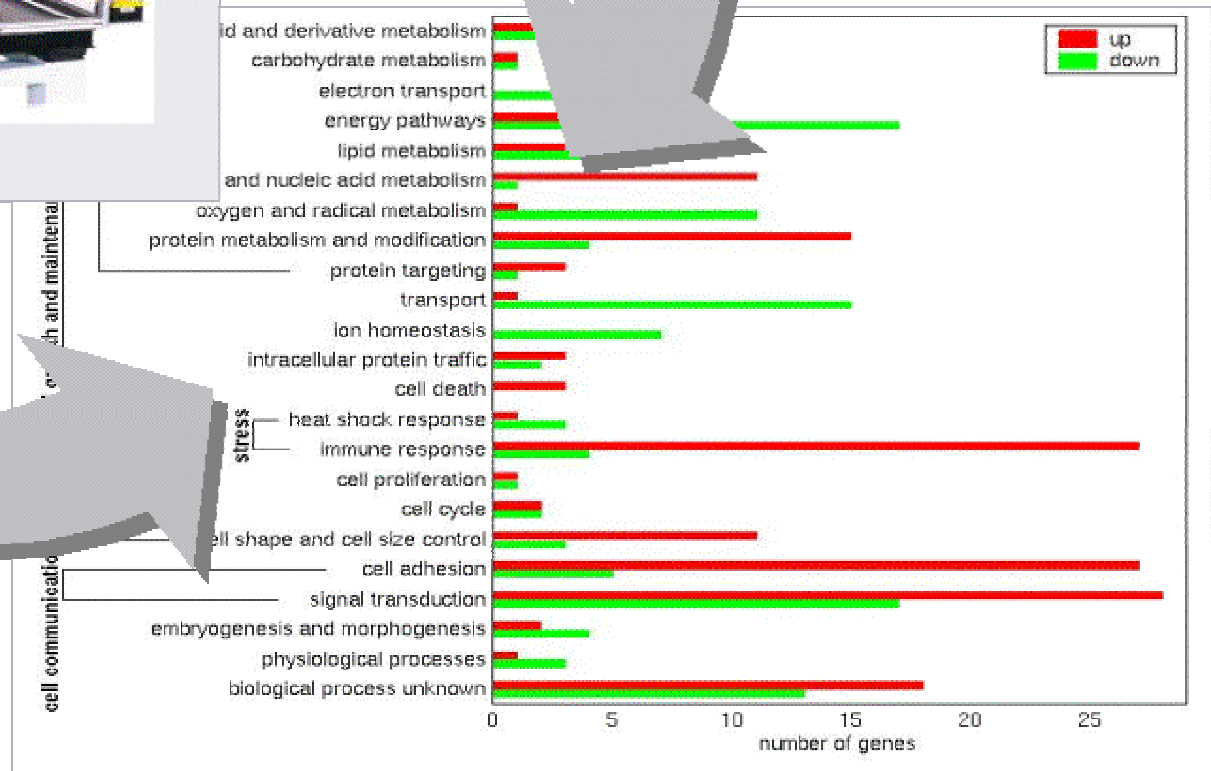
integration of high-throughput array technologies



**protein
arrays**



**cDNA
arrays**



RZPD Clone ID	Length	ORF Start	ORF Stop	Map Position	Similarity	Protein Group	Localization	Kidney	Stroma	Brain	Breast
DKFZp434E248	3290	21	1994	3q29	similar to: putative GTP-binding protein	signaling & communication	Endoplasmic Reticulum				
DKFZp434G058	2762	105	1736	12q24.22	unknown	unknown	Cytosol				
DKFZp434J0460	1180	38	660	17q26.1	unknown	unknown	Nucleus				
DKFZp434M108	2107	45	1937	8q22.2	similar to: poly(A) binding protein	nucleic acid management	Cytosol				
DKFZp434N214	3010	466	2429	9q34.11	similar to: RNA-binding protein/RNP	nucleic acid management	Other/unknown				
DKFZp434P097	3003	67	2120	12q13.13	transcription control	nucleic acid management	Cytosol				
DKFZp434P2235	2027	99	1745	17q12	similar to: oncogene	cell cycle	Cytosol + Nucleus				
DKFZp564A202	707	33	620	10q24.31	similar to: protein involved in energy metabolism	metabolism	Endoplasmic Reticulum				
DKFZp564B0482	2092	317	1579	17q11.2	unknown, contains 2 WVD-40 repeats, which are typical for the beta-transducin subunit of G-proteins	signaling & communication	Cytosol + Nucleus				
DKFZp564B1023	2905	157	1890	1q23.2	similar to: RNA helicase	nucleic acid management	Nucleus				
DKFZp564B1162	4503	661	2626	4q22.1	unknown	signaling & communication	Cytoskeleton				
DKFZp564B2123	3300	121	699	8q22.2	similar to: Neuro calcin	signaling & communication	Cytosol + Nucleus				
DKFZp564B246	2054	73	1074	5q34	similar to: dTDP-6-deoxy-L-mannose-dehydrogenase	metabolism	Nucleus				
DKFZp564C162	2306	166	626	1p36.11	unknown	membrane protein	Endoplasmic Reticulum				
DKFZp564D116	2535	29	1840	10q24.1	similar to: GTP-binding protein	signaling & communication	Cytosol				
DKFZp564D152	3229	99	1412	8q22.1	unknown	differentiation & development	Cytosol + Nucleus				
DKFZp564E1782	1818	40	972	14q23.1	unknown	membrane protein	Endoplasmic Reticulum				
DKFZp564E2182	2367	193	804	6q22.1-q22.33	similar to: protein involved in cell cycle, DNA repair, maintenance of minichromosomes	Cell Cycle	Nucleus				
DKFZp564H012	957	93	632	6q26	unknown	unknown	Mitochondria				
DKFZp564H1122	1734	159	1133	11q14	unknown	membrane protein	Nucleus				
DKFZp564M082	902	227	589	Xp22	unknown, contains osteopontin motif	unknown	Cytosol + Nucleus				
DKFZp566I1024	1783	5	970	7p11.2	similar to: hypothetical protein Rv0712 - Mycobacterium tuberculosis	unknown	Cytosol				
DKFZp566K144	3084	466	1232	11q13.4	similar to: Rab protein	transport and traffic	Golgi				
DKFZp566B0719	1640	108	1258	11q14.2	similar to: serine protease	signaling & communication	Nucleus				
DKFZp566E1323	1854	367	954	12q24.23	similar to: heat shock protein	protein management	Cytosol + Nucleus				
DKFZp566I0418	1568	163	822	7q31	unknown	unknown	Cytosol + Nucleus				
DKFZp566J1023	1048	72	748	7q34	similar to: glutathione S-transferase / posttranslational modification	protein management	Cytosol + Nucleus				
DKFZp566P2421	1995	69	902	9q22.2	similar to: corneal keratan sulfate proteoglycan sulfate proteoglycan	differentiation & development	Endoplasmic Reticulum				
DKFZp761P041	3578	19	1874	19p13.3	unknown	unknown	Endoplasmic Reticulum				

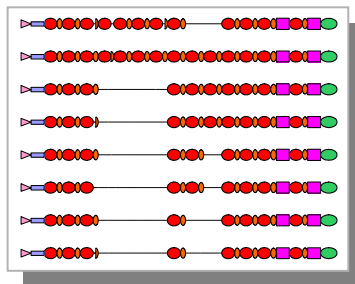
RZPD Clone ID	Length	ORF Start	ORF Stop	Map Position	Similarity	Protein Group	Localization	Kidney	Stromal	Brain	Breast
DKFZp434E248	3280	21	1994	3q29	similar to: putative GTP-binding protein	signaling & communication	Endoplasmic Reticulum	█			
DKFZp434G058	2762	105	1736	12q24.22	unknown	unknown	Cytosol	█			
DKFZp434J0460	1180	36	660	17q26.1	unknown	unknown	Nucleus				█
DKFZp434M108	2107	46	1837	8q22.2	similar to: poly(A) binding protein	nucleic acid management	Cytosol	█			
DKFZp434N214	3010	466	2428	9q34.11	similar to: RNA-binding protein/RNP	nucleic acid management	Other/Unknown	█			
DKFZp434P087	3003	67	2120	12q13.13	transcription control	nucleic acid management	Cytosol				█
DKFZp434P2236	2027	99	1746	17q12	similar to: oncogene	cell cycle	Cytosol + Nucleus	█			
DKFZp434Q202	707	33	620	10q24.31	similar to: p:protein involved in energy metabolism	metabolism	Endoplasmic Reticulum				█
DKFZp434S482	2092	317	1579	17q11.2	unknown, contains 2 WD-40 repeats, which are typical for the base-translocin subunit of G-proteins	signaling & communication	Cytosol + Nucleus	█			
DKFZp434T23	2909	157	1990	1q23.2	similar to: RNA helicase	nucleic acid management	Nucleus				█
DKFZp434B1182	4893	661	2626	4q22.1	unknown	signaling & communication	Cytoskeleton	█			
DKFZp434B2123	3300	121	699	8q22.2	similar to: Neurocalcin	signaling & communication	Cytosol + Nucleus	█			
DKFZp434B246	2054	73	1074	9q3.4	similar to: d TDP- β -D-glucose-6-phosphate dehydrogenase	metabolism	Nucleus				█
DKFZp434C162	2306	166	626	1p38.11	unknown	membrane protein	Endoplasmic Reticulum				█
DKFZp434D116	2636	29	1840	10q24.1	similar to: GTP-binding protein	signaling & communication	Cytosol				█
DKFZp434D152	3229	99	1412	6q22.1	unknown	diff. renitiation & development	Cytosol + Nucleus				█
DKFZp434E1782	1818	40	972	14q23.1	unknown	membrane protein	Endoplasmic Reticulum				█
DKFZp434E2182	2367	193	804	8q22.1-q22.33	similar to: p:protein involved in cell cycle, DNA re pair, maintenance of mitochondriosomes	Cell Cycle	Nucleus				█
DKFZp434H012	957	93	632	6q26	unknown	unknown	Mitochondria				█
DKFZp434H1122	1734	159	1133	11q14	unknown	membrane protein	Nucleus	█			
DKFZp434M082	902	227	669	3p22	unknown, contains osteonin motif	unknown	Cytosol + Nucleus	█			
DKFZp434I1024	1763	5	970	7p11.2	similar to: hypothetical protein R40712 - Mycobacterium tuberculosis	unknown	Cytosol				█
DKFZp434K144	3084	466	1232	11q13.4	similar to: Rab protein	transport and trafio	Golgi	█			
DKFZp434B0719	1640	108	1266	11q14.2	similar to: serine protease	signaling & communication	Nucleus	█			
DKFZp434E1323	1954	367	954	12q24.23	similar to: heat shock protein	protein management	Cytosol + Nucleus	█			
DKFZp434B10418	1668	163	622	7q31	unknown	unknown	Cytosol + Nucleus				█
DKFZp434B11023	1040	72	740	7q34	similar to: glutathione S-transferase / posttranslational modification	protein management	Cytosol + Nucleus				█
DKFZp434B2421	1995	69	902	9q22.2	similar to: corneal keratan sulfate proteoglycan sulfate proteoglycan	diff. renitiation & development	Endoplasmic Reticulum				█
DKFZp434B19041	3578	19	1674	19p13.3	unknown	unknown	Endoplasmic Reticulum	█			

redundancy helps

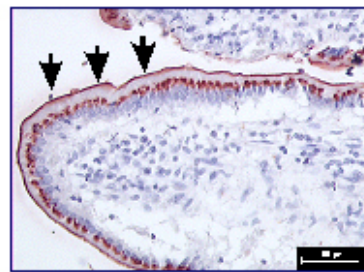
DKFZp564H1122	DKFZp434M108
tumor stage I III IV M	tumor stage I III IV M
UniGene clusters IMAGp950D0637	UniGene clusters IMAGp950A2250
IMAGp950A1532	IMAGp950C1919
IMAGp950B2437	IMAGp950B0949
IMAGp950J1631	

complex systems require complex approaches

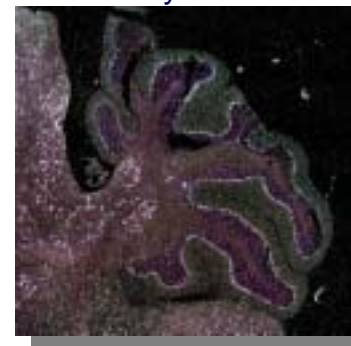
domain structure



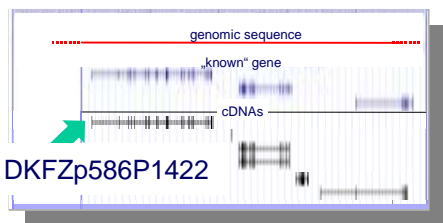
immuno-histochemistry



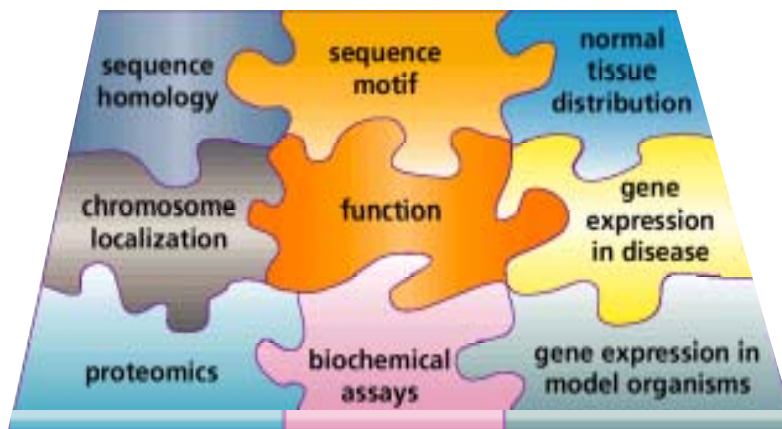
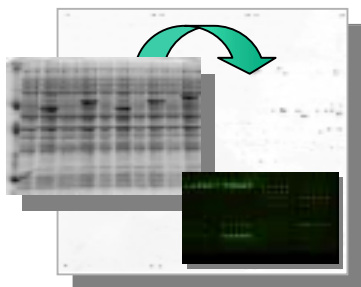
in situ hybridization



gene structure

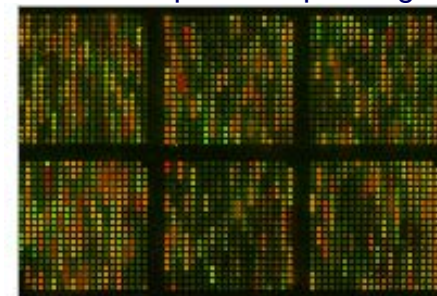


peptide/compound arrays

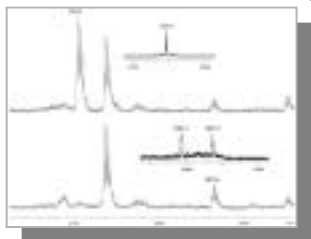


C Debouck & P N Goodfellow, *Nature Genet.* 21, 48 – 50 (1999)

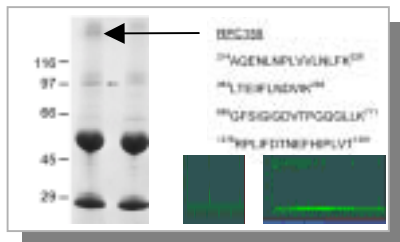
RNA expression profiling



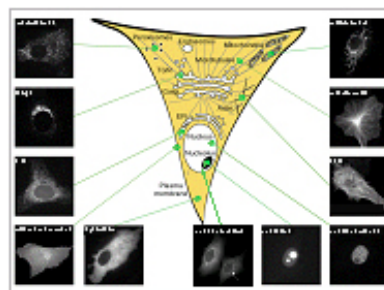
post-translational modification



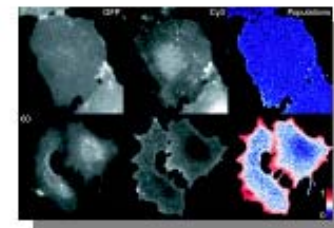
protein-protein interaction



protein localization

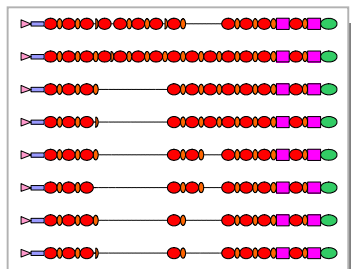


functional assays

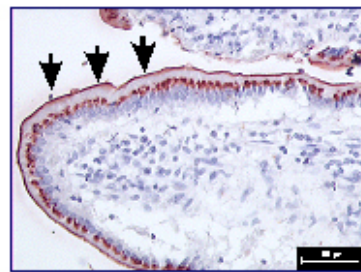


complex systems require complex approaches

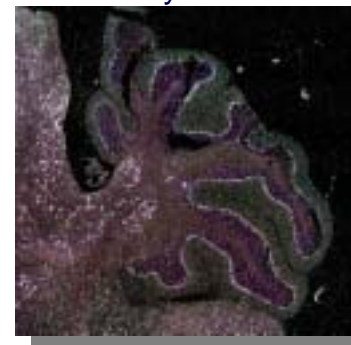
domain structure



immuno-histochemistry



in situ hybridization

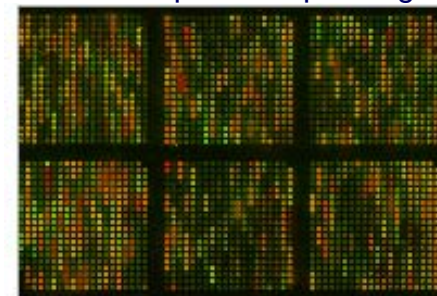


gene structure

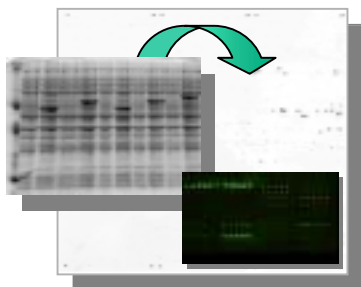


... what is „function“ ?

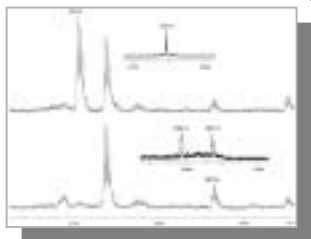
RNA expression profiling



peptide/compound arrays



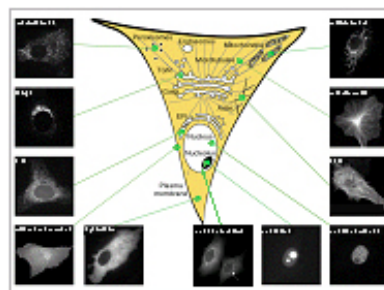
post-translational modification



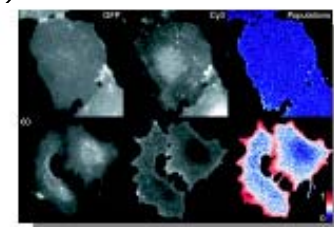
protein-protein interaction



protein localization



functional assays



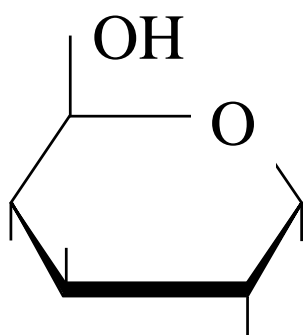
what is „function“ ?

example: enzyme reaction (catalysis)

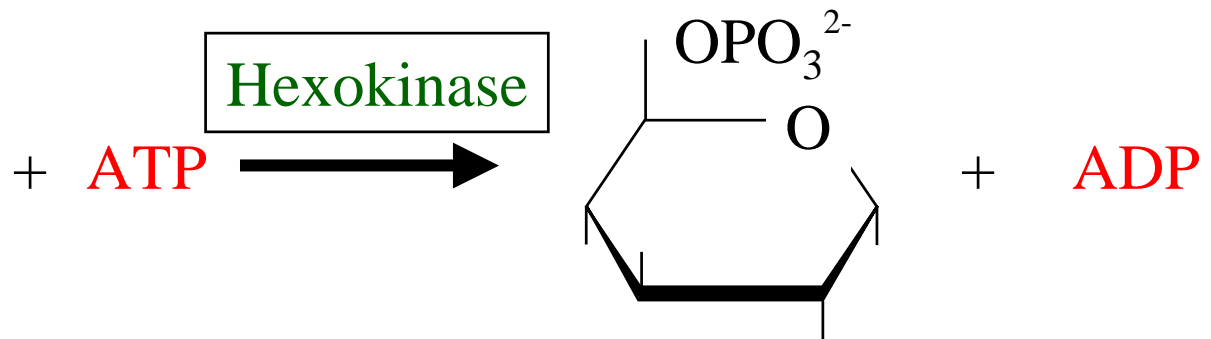


Educts

Products

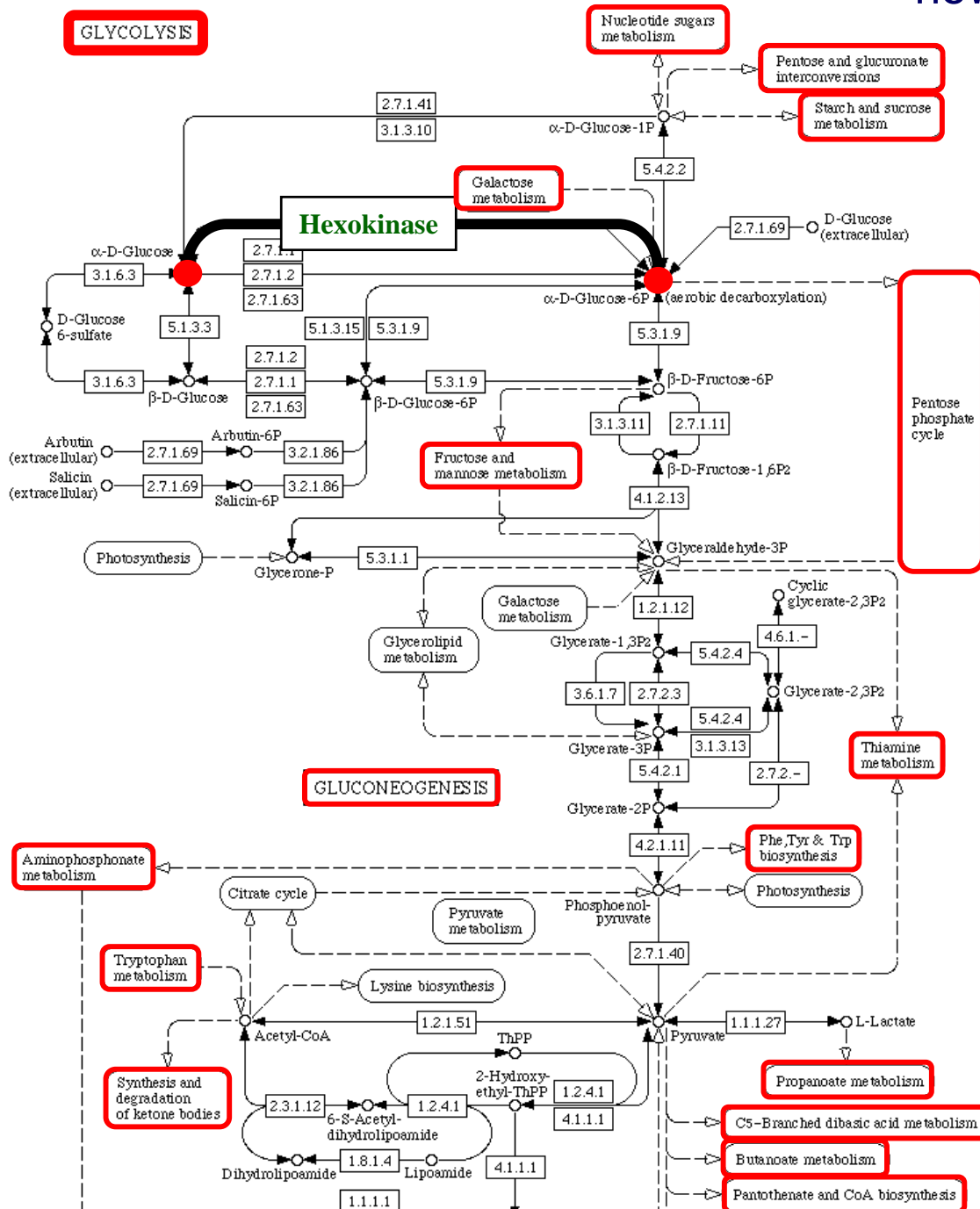


Glucose

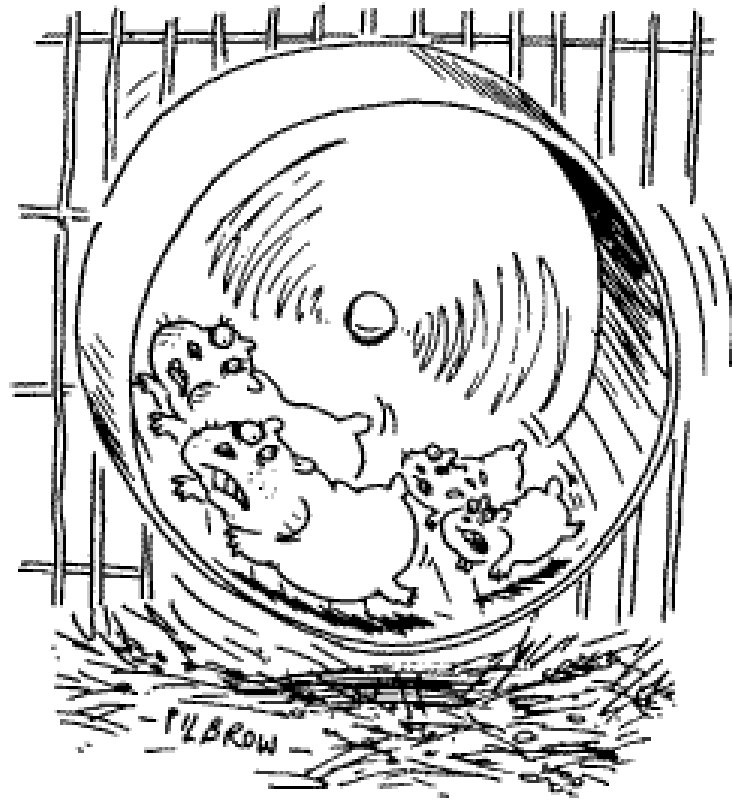


Glucose-6-phosphate

how many dimensions are needed to describe a gene/protein ?



1. catalytic activity
2. (metabolic) pathway
3. direct interplay with other pathways
(e.g. Glycolysis - Gluconeogenesis)
4. variants of a gene
(on RNA level - splicing)
5. variants of a protein
(posttranslational modifications)
6. where in the cell / in the tissue / in the organism is the protein localized
(possible interaction partners)
7. when is the protein expressed
(development / differentiation)
8. ... **the "Omeome"**
(try to top this ☺)



"Are we there yet, daddy?"

Taken from The History of Private Eye Cartoons

DKFZ Heidelberg
Molecular Genome Analysis
Annemarie Poustka

Ruth Wellenreuther	libraries
Regina Albert	sequencing
Ute Ernst	Taqman
Birgit Guilleaume	proteomics
Ulrike Korf	proteomics
Steffi Bechtel	Gateway cloning
Dorit Arlt	cell biology
Lee Bergman	cell biology
Mamatha Sabbela	cell biology
Detlev Bannasch	databases
Vladimir Kurshev	data integration
Holger Sültmann	expression profiling
Anja Kolb	<i>in situ</i> hyb

Protein Analytics Facility
Martina Schnölzer

University of Göttingen
Laszlo Fuezesi

The German cDNA Consortium

S. Wiemann	DKFZ Heidelberg
W. Ansorge	EMBL Heidelberg
H. Blöcker	GBF Braunschweig
J. Lauber	QIAGEN Hilden
K. Köhrer	BMFZ Düsseldorf
W. Mewes	MIPS München
B. Ottenwälder	MediGenomix München
R. Wambutt	AGOWA Berlin
A. Poustka	DKFZ Heidelberg

EMBL

Rainer Pepperkok
Jeremy Simpson
Philippe Bastiaens

Resource Center
Bernhard Korn

and ...

- Robert Strausberg
Mammalian Gene Collection
- Sumio Sugano
NEDO project
- Osamu Ohara
Kazusa DNA Research Center
- providers of
cDNA + genomic sequences in
public databases
and matching clones in
clone repositories

