

# V10 Protein-Protein-Interaktionsnetzwerke

Rückblick (V7): Arten von PP-Interaktionen:

## **Homo-Oligomere vs. Hetero-Oligomere**

Homo-Oligomere bestehen aus mehreren identischen Einheiten und werden z.B. von Ionen-Kanälen oder Rezeptoren gebildet

## **Stabile vs. Transiente Komplexe**

Stabil: Ribosom, RNA-Polymerase, ...

Transient: Redox-Partner, Signaltransduktion

## **Obligate vs. Nicht-obligate Komplexe:**

obligat (-> obligatorisch): Komponenten liegen in der Zelle nur als Komplex vor

nicht-obligat: Komponenten existieren in der Zelle ebenfalls im freien Zustand (z.B. Antikörper)

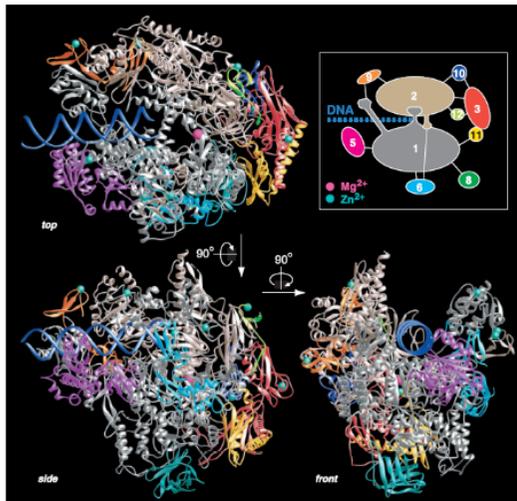
## **Kovalente vs. Nicht-kovalente Komplexe**

Kovalente: z.B. Ubiquitin-modifizierte Proteine

Nicht-kovalent: ist der übliche Fall

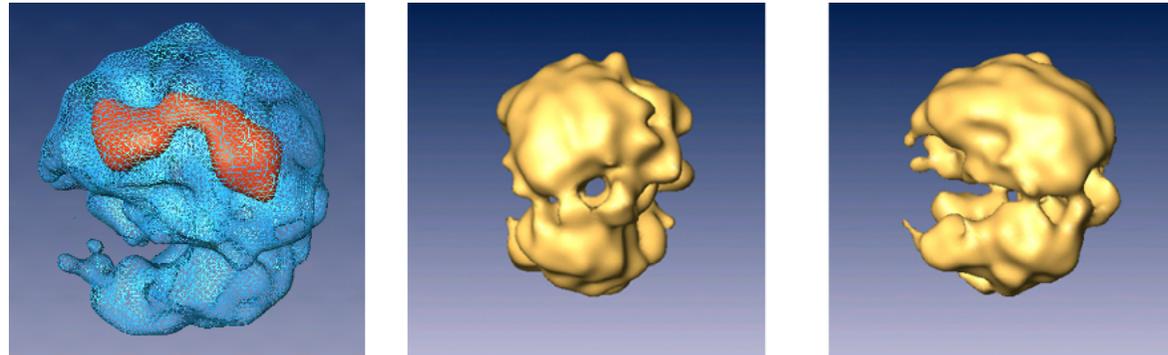
# Beispiele für wichtige Proteinkomplexe

RNA Polymerase II



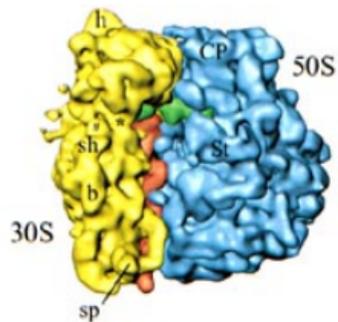
Cramer et al., Science 288, 640 (2000)

Spliceosome



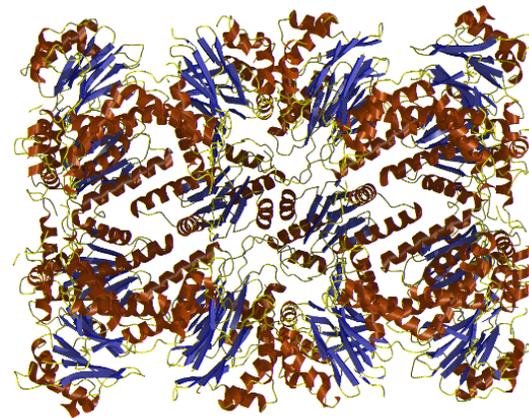
<http://www.weizmann.ac.il/>

Ribosom



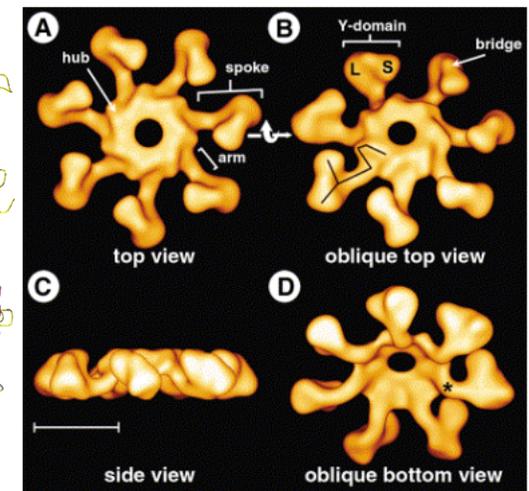
<http://www.millerandlevine.com/chapter/12/cryo-em.html>

Proteasome



<http://www.biochem.mpg.de>

Apoptosom



Acehan et al. Mol. Cell 9, 423 (2002)

# Proteininteraktionsnetzwerke

The screenshot displays the Cytoscape Desktop interface (Session: humanANDmouseIntact.cys) with several panels:

- Control Panel:** Includes tabs for Network, VizMapper™, Editor, Filters, and MCODE. A table shows network statistics:

Network	Nodes	Edges
Mouse Interactome from IntAct	3007(1)	5825(136)
Human Interactome from IntAct	8578(52)	30491(817)
MCODE Result 1 - Cluster 1	4(0)	69(0)
MCODE Result 1 - Cluster 3	13(0)	161(0)
MCODE Result 1 - Cluster 1	52(0)	406(0)
- Network View:** A large network graph with nodes and edges. A zoomed-in view shows a cluster of nodes with labels Q00987, Q93009, and P04637.
- Search Panel:** Shows search results for "TP53".

Step 1: Search: TP53 (All Organisms)

Examples: TP53, MGA1, or SKY.

Step 2: Select: TP53, TP53 activated protein 1, TP53 regulating kinase, TP53 regulating kinase.

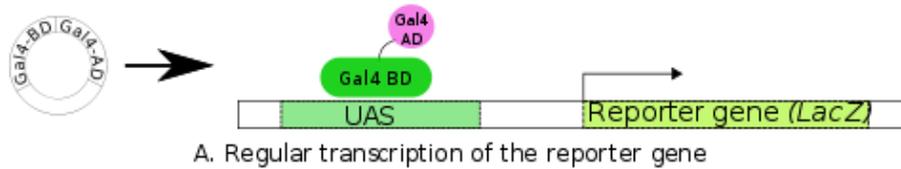
Step 3: Select Network(s): Pathways (p53-independent G..., G2/M Checkpoints, G2/M DNA damage c..., DNA Repair, Double-Strand Break..., Homologous Recomb..., Homologous recomb..., ATM mediated respo..., Cell Cycle, Mitotic, Signaling mediated b..., Role of HDAC Class I) and Interaction Networks (Reactome).
- Results Panel:** Displays MCODE results for Cluster 1, including Rank, Score, Nodes, and Edges.
- Data Panel:** Shows node details for ID P62158, MI [0499], and ensembl [ENSG00000143933, ENSG00000160014, EN...].

At the bottom, it says "Welcome to Cytoscape 2.6.0" and provides navigation instructions: "Right-click + drag to ZOOM" and "Middle-click + drag to PAN".

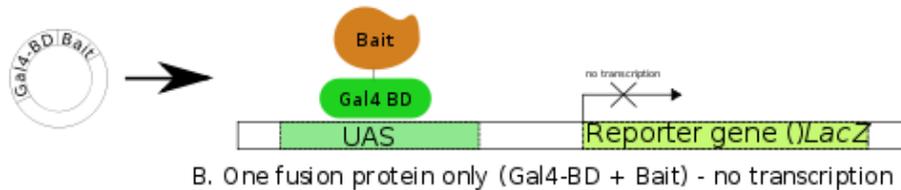
Screenshots

# Detektiere PP-Interaktionen: Yeast Two-Hybrid Methode

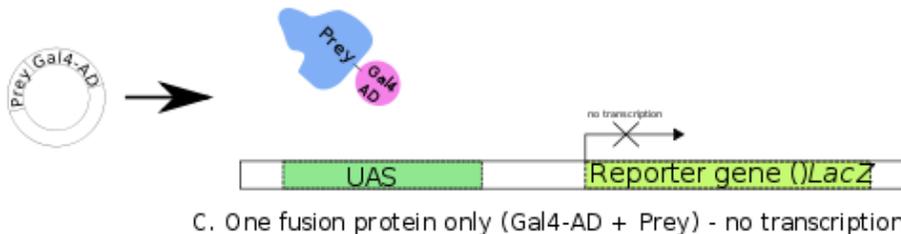
Ziel: entdecke binäre PPIs zwischen einem "bait" Protein (*dt. Köder*) und einem "prey" Protein (*dt. Beutetier*), die "physikalisch", d.h. direkt miteinander wechselwirken.



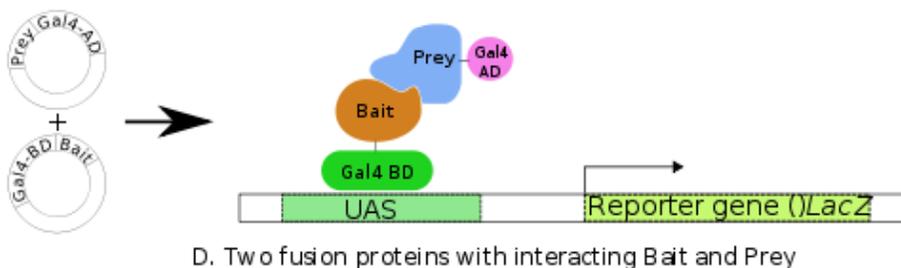
Gegeben: Transkriptionsfaktor, der ein Reporter-Gen reguliert, besteht aus einer DNA-bindenden Domäne (BD) und einer Aktivator-domäne (AD)



Unterbreche kovalente Verbindung BD-AD; verbinde bait (orange) mit BD und prey (grün) mit AD → Expression findet nur statt, wenn bait:prey-Komplex gebildet wird



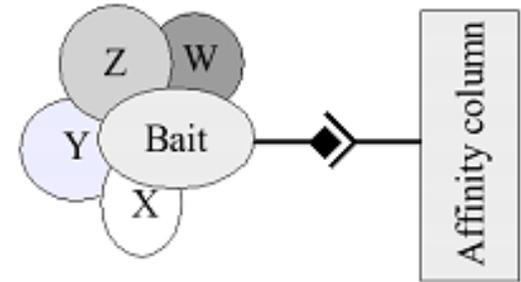
Man kann das normale Reporter-Gen (hier lacZ) auch durch ein GFP-Gen ersetzen. Dann kann man erfolgte Transkription, d.h. Bindung von bait:prey-Komplex, als Fluoreszenz detektieren.



# Tandem affinity purification (also „pull-down“)

Yeast 2-Hybrid-Methode kann nur binäre Komplexe identifizieren.

In der **Affinitäts-Aufreinigung** wird ein bestimmtes Protein (bait) mit einem molekularen Label verbunden (dunkle Route in Abb.) um dessen Aufreinigung zu erleichtern. Das so “ge-taggte” Protein wird dann in einer Affinitätssäule aus der Lösung herausgefischt, zusammen mit allen interagierenden Partnern (W, X, Y, Z).

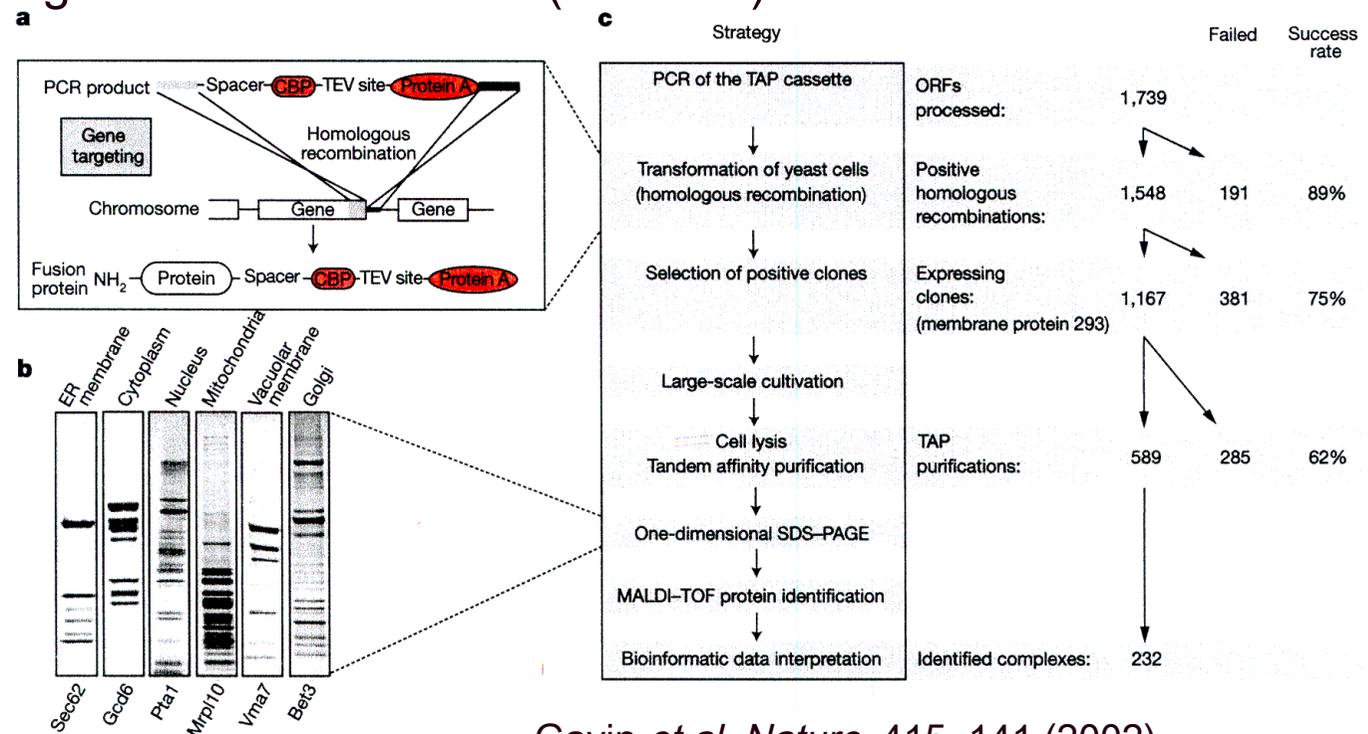


Diese Strategie kann man genomweit anwenden (wie Y2H).

Hier gezeigt:  
Anwendung für  
*S. cerevisiae*.

Identifiziere Proteine in  
Gelbanden mittels  
Massenspektrometrie.

Label unten: bait-Protein  
Jede Bahn entspricht einem  
Komplex

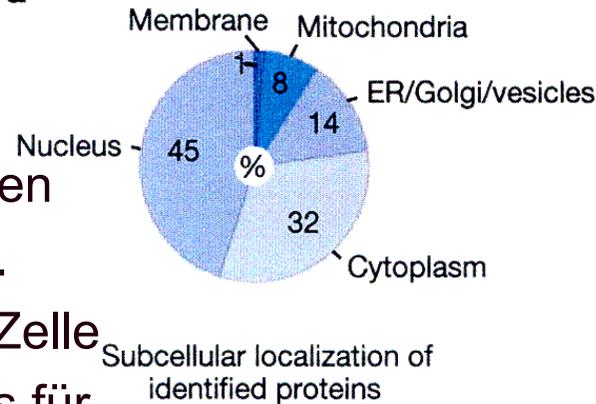


Gavin et al. Nature 415, 141 (2002)

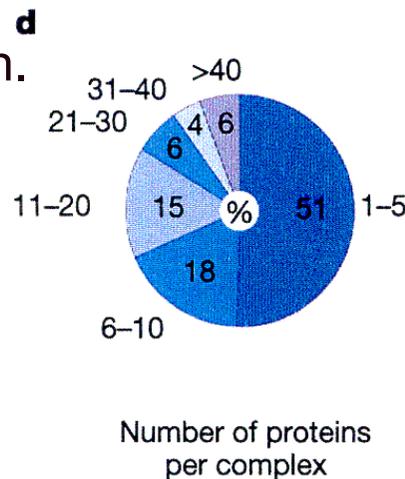
# TAP-Analyse für PP-Komplexe in *S. cerevisiae*

Identifiziere Proteine durch die Massen ihrer Peptidfragmente.

(a) listet die an Komplexen beteiligten Proteine bzgl. ihrer Lokalisation in der Zelle -> es scheint keinen Bias für bestimmte Kompartments zu geben.

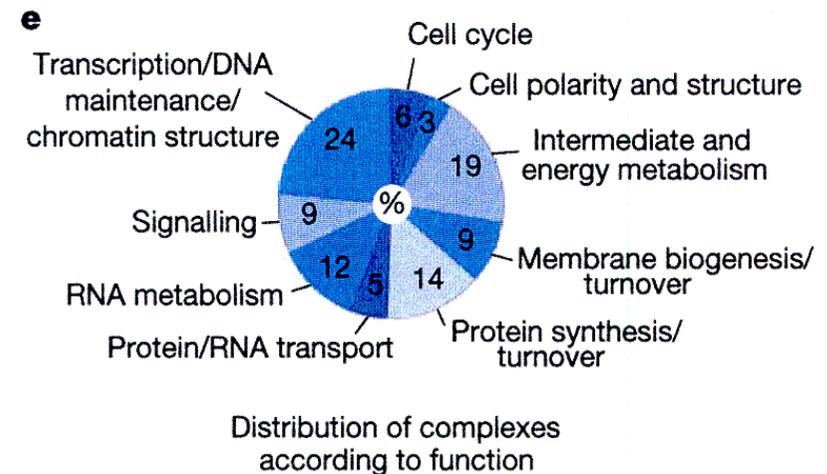


Allerdings findet man nur wenig Membranproteine (Anteil sollte ca. 25% sein)



(d) die Hälfte aller PP-Komplexe hat 1-5 Mitglieder, die andere Hälfte ist grösser.

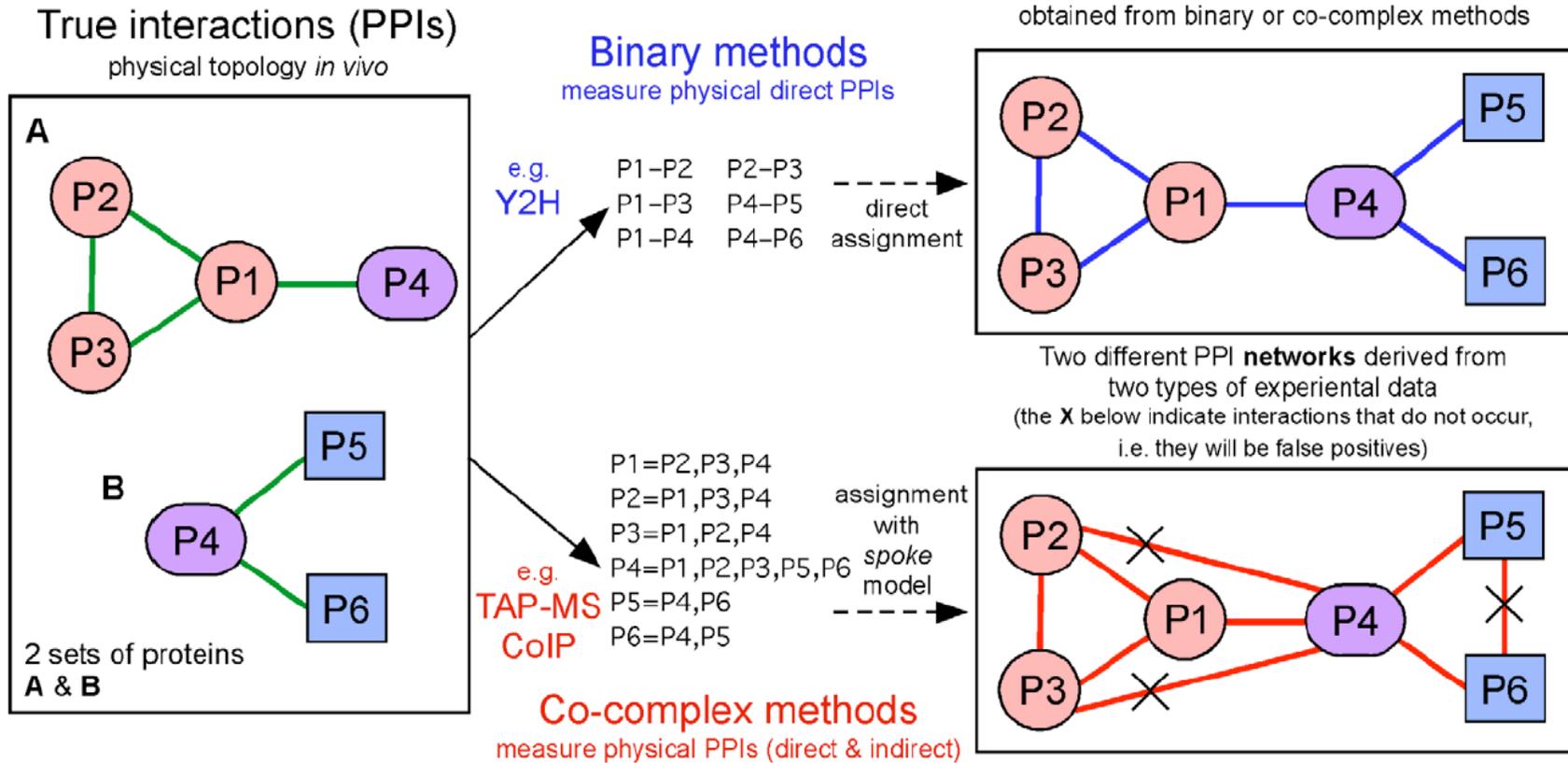
(e) PP-Komplexe sind an praktisch allen zellulären Prozessen beteiligt.



Gavin *et al.* *Nature* 415, 141 (2002)

# Protein-Interaktionsnetzwerke

Unterschiedliche experimentelle Techniken messen verschiedene Eigenschaften von Proteinkomplexen. Das sind keine Messfehler, sondern Eigenheiten der einzelnen Methoden.



In der Probe gibt es zwei Arten von Komplexen, A und B

Ergebnis der Messung für diese Probe.

„Logische“ Interpretation der Messergebnisse.

De Las Rivas, PLOS Comp Biol. 6, e1000807 (2010)

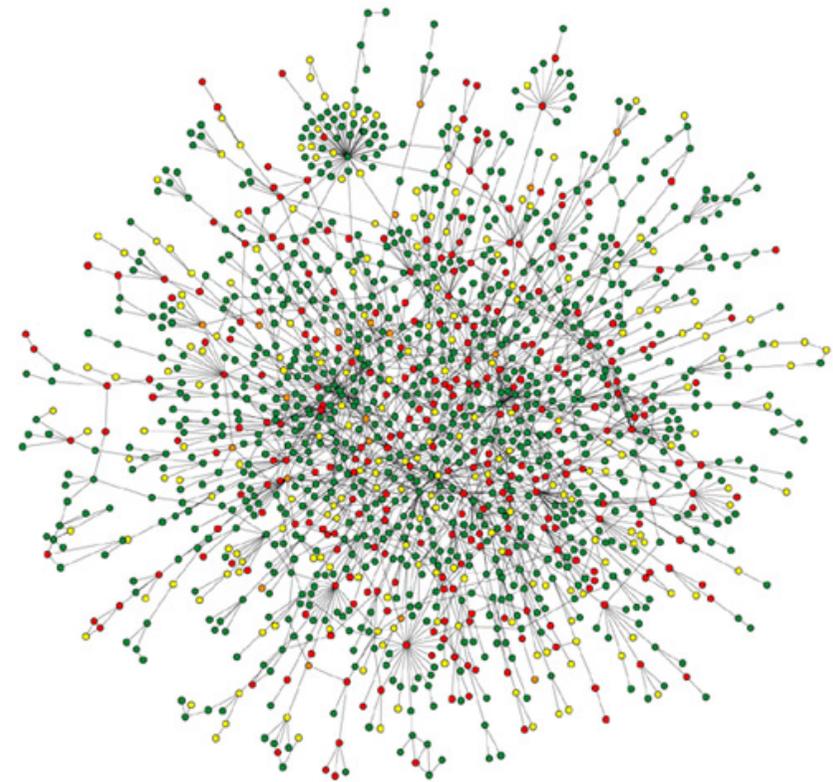
# Globales Protein-Interaktionsnetzwerk in *S. cerevisiae*

Abb. zeigt das Protein–Protein Interaktionsnetzwerk in *Saccharomyces cerevisiae*, basierend auf **yeast two-hybrid** Experimenten.

Jeder Knoten (Kreis): bestimmtes Protein  
Verbindungen/Kanten: physikalische Interaktionen.

Das Netzwerk wird von einigen Knoten mit sehr vielen Verbindungen zusammengehalten. Diese nennt man **Hubs**.

Dieser größte Cluster enthält 78% aller Proteine der Zelle.



Nature Reviews | Genetics

**Rot:** Gendeletion ist tödlich

**Grün:** Gendeletion ist nicht tödlich

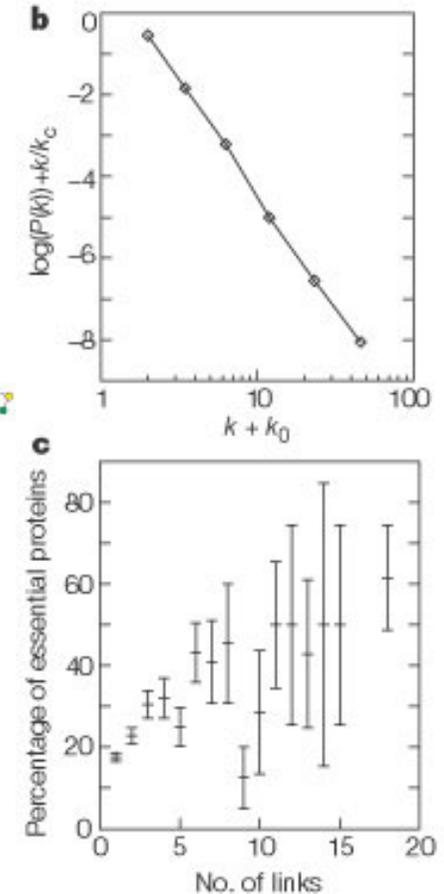
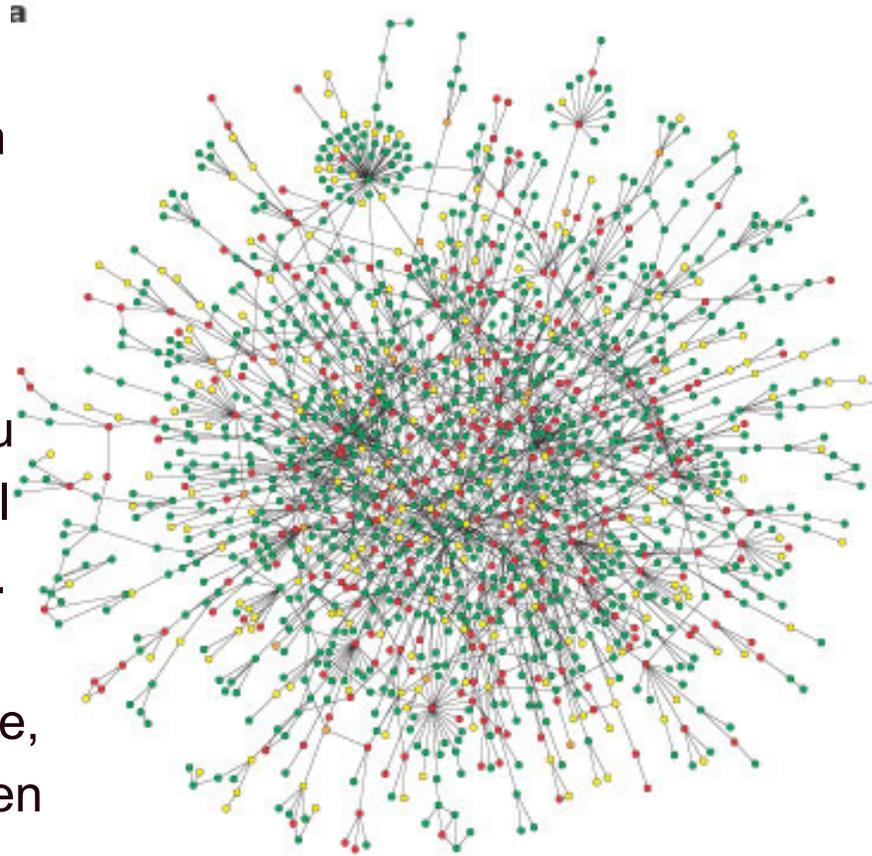
**Gelb:** Effekt der Gendeletion ist unbekannt.

Barabasi & Oltvai, Nature Rev Gen 5, 101 (2004)

# Welche Proteindeletionen sind tödlich?

(b) Die Häufigkeit von Hub-Proteinen mit  $k$  Interaktionen nimmt nicht exponentiell ab (wie man dies in einem Zufallsnetzwerk erwartet), sondern etwa proportional zu  $1 / k^2$  oder  $1 / k^3$ , ist also viel häufiger als zufällig erwartet.

(c) Die Deletion solcher Gene, die für Hub-Proteine mit vielen Links kodieren (und im Plot rechts liegen), ist eher **tödlich** für die Zelle als die Deletion von Proteinen, die mit wenigen anderen Proteinen interagieren (links im Plot).



„essentielle“ Proteine: Zelle ist nicht lebensfähig ohne diese Proteine

H. Jeong, S. P. Mason, A.-L. Barabási and Z. N. Oltvai  
*Nature* **411**, 41-42

# Wieviele Proteininteraktionen gibt es

## *S. cerevisiae*

BioGrid ([www.thebiogrid.org](http://www.thebiogrid.org)): 91,651 nicht-redundante physikalische Interaktionen von 6367 Hefe-Proteinen (August 2017).

Mentha (<http://mentha.uniroma2.it/>): 106,683 Interaktionen.

PrePPI (<https://bhapp.c2b2.columbia.edu/PrePPI/>).  
(bioinformatisch integrativ kompilierter Datensatz):  
60000 Interaktionen mit hoher Zuverlässigkeit

## Mensch

Mentha: 277,371 physikalische Interaktionen zwischen 18,506 menschlichen Proteinen.

PrePPI: 1.35 Millionen vorhergesagte Interaktionen.  
Für 127,000 davon gibt es experimentelle Bestätigung der direkten Interaktion.

# In-Silico Vorhersagemethode

## **Sequenz**-basiert:

- Gen-clustering
- Gen-Nachbarschaft
- Rosetta stone
- phylogenetisches Profiling
- Ko-Evolution

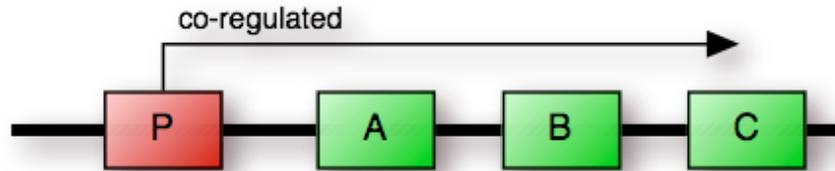


## **Struktur**-basiert:

- interface propensities (V7)
- Protein-Protein Docking (V7)
- 3D-Simulationen (z.B. MD)

# Gen-Clustering

**Idee:** funktionell **verwandte** Proteine oder Teile eines Komplexes werden **gleichzeitig** exprimiert



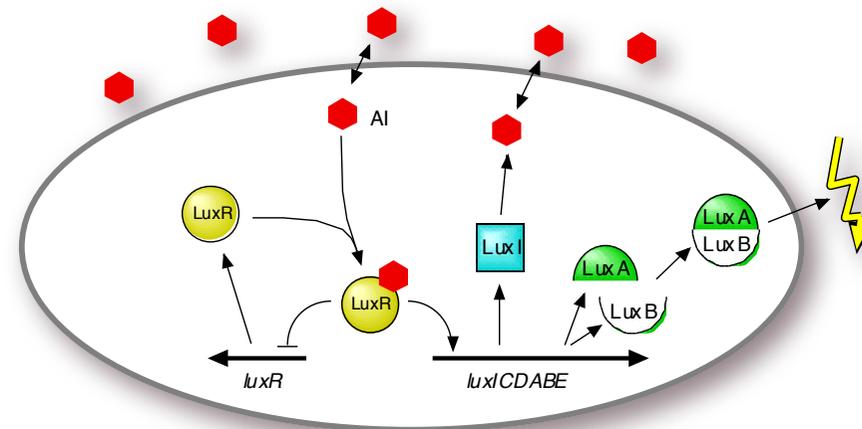
Suche nach Genen mit einem **gemeinsamen Promoter**

→ wenn aktiviert, werden alle gemeinsam als ein *Operon* transkribiert.

## Beispiel:

Biolumineszenz in *V. fischeri* wird durch Quorum sensing reguliert  
→ 3 Proteine I, AB, CDE sind dafür verantwortlich.

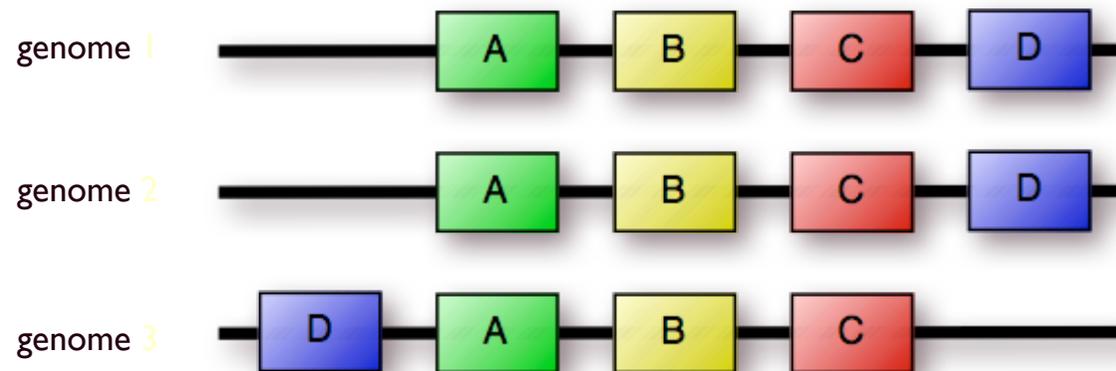
Sie sind als 1 Operon namens <sup>12</sup>*luxICDABE* organisiert.



# Gen-Nachbarschaft

**Hypothese:** funktionell **verwandte** Gene werden **gemeinsam exprimiert**

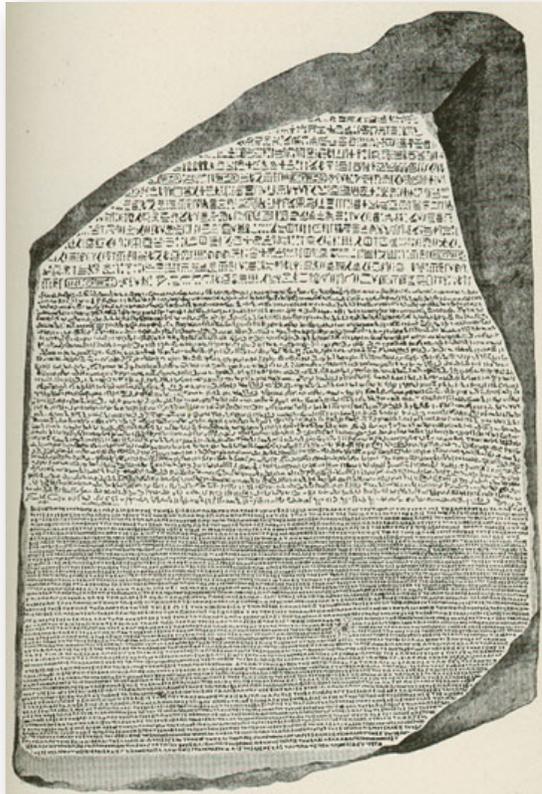
"funktionell verwandt" heißt gleicher {Komplex | Pfad | Funktion | ...}



→ Suche nach **ähnlicher Anordnung** der verwandten Gene in **verschiedenen Organismen**

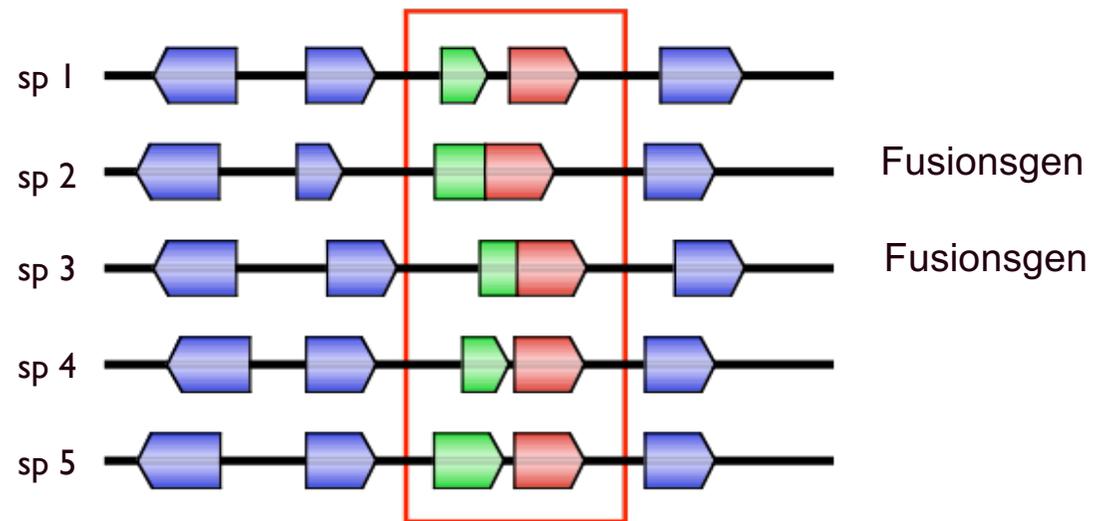
( $\Leftrightarrow$ ) Gen-Clustering: in einem Organismus, Promoter müssen bekannt sein

# Rosetta Stein Methode



Mehrsprachige Stele aus 196 v.Chr.,  
wurde 1799 gefunden.  
Auf dem Stein steht derselbe Text in 3  
Sprachen: Hieroglyphen, demotische  
Schrift, griechisch  
→ Schlüssel um Bedeutung der  
Hieroglyphen zu entschlüsseln

**Idee:** finde homologe Gene ("**Worte**") in den Genomen  
verschiedener Organismen ("**Texte**")  
- Überprüfe, ob ein Organismus ein Fusions-Gen enthält  
→ Kann darauf hindeuten, dass die beiden Proteine einen  
Komplex bilden

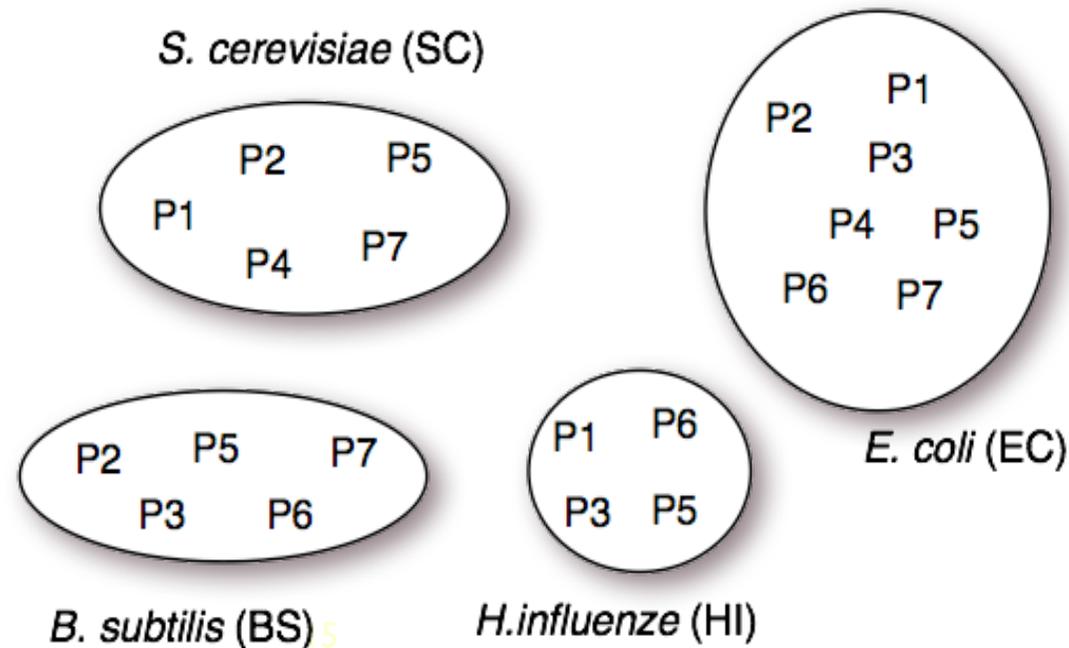


Enright, Ouzounis (2001):  
40000 vorhergesagte paarweise Interaktionen  
in 23 Spezies

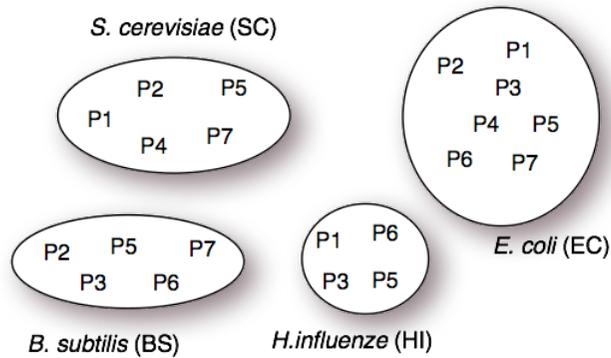
# Phylogenetisches Profiling

**Idee:** entweder **alle** oder **kein** Protein eines Komplexes sollten in einem Organismus vorkommen

→ Vergleiche Vorkommen homologer Proteine zwischen Spezies  
(z.B. mit Sequenzalignment)



# Distanzen in Phylogenetischem Profiling

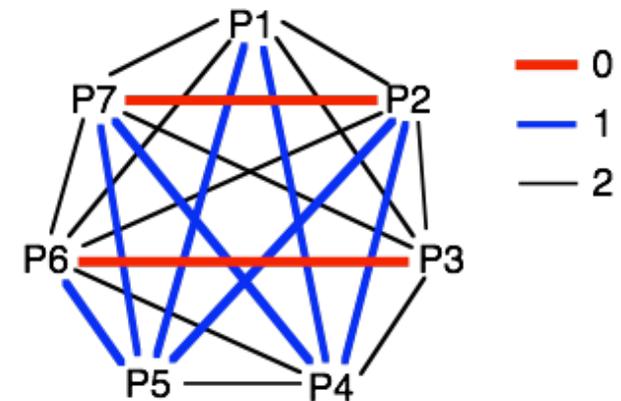


Kodierte Vorkommen/Abwesenheit

	EC	SC	BS	HI
P1			0	
P2				0
P3		0		
P4			0	0
P5				
P6		0		
P7				0

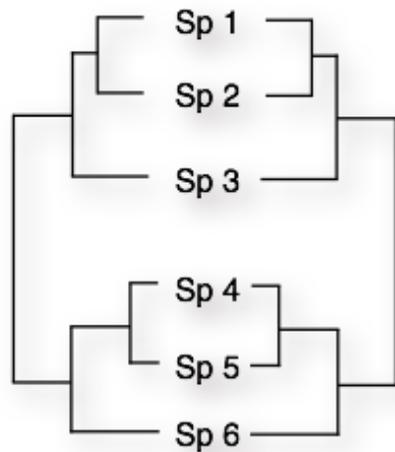
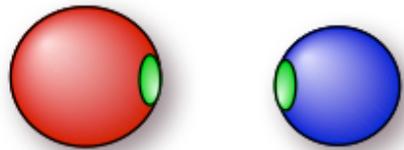
**Hamming-Distanz** zwischen Spezies: Anzahl an unterschiedlichem Vorkommen

	P1	P2	P3	P4	P5	P6	P7
P1	0	2	2			2	2
P2		0	2			2	<b>0</b>
P3			0	3		<b>0</b>	2
P4				0	2	3	
P5					0		
P6						0	2
P7							0



Paare mit ähnlichem Vorkommen sind: P2-P7 und P3-P6

Dies sind Kandidaten für Protein-Interaktionen.



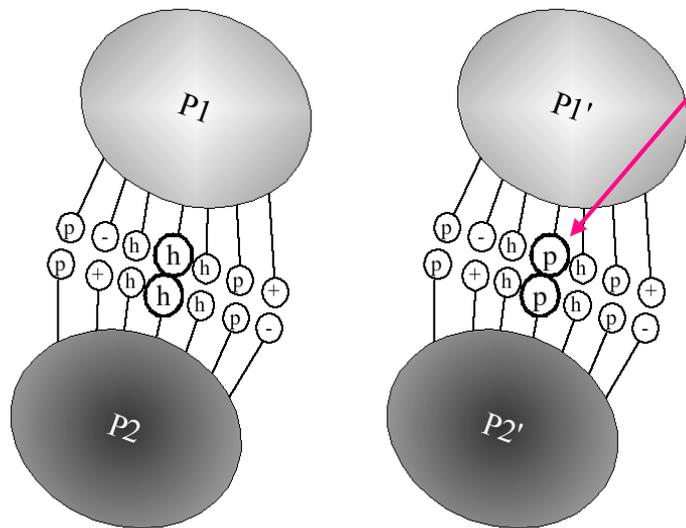
# Ko-Evolution

Bindungsschnittstellen von Komplexen sind nur **leicht stärker konserviert** als die restliche Proteinoberfläche.

**Idee** von Pazos & Valencia (1997):

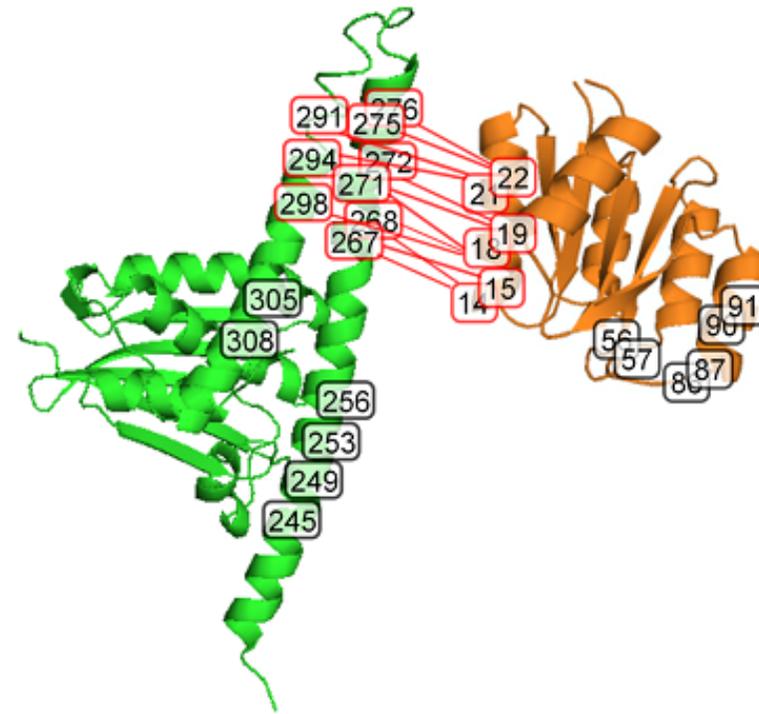
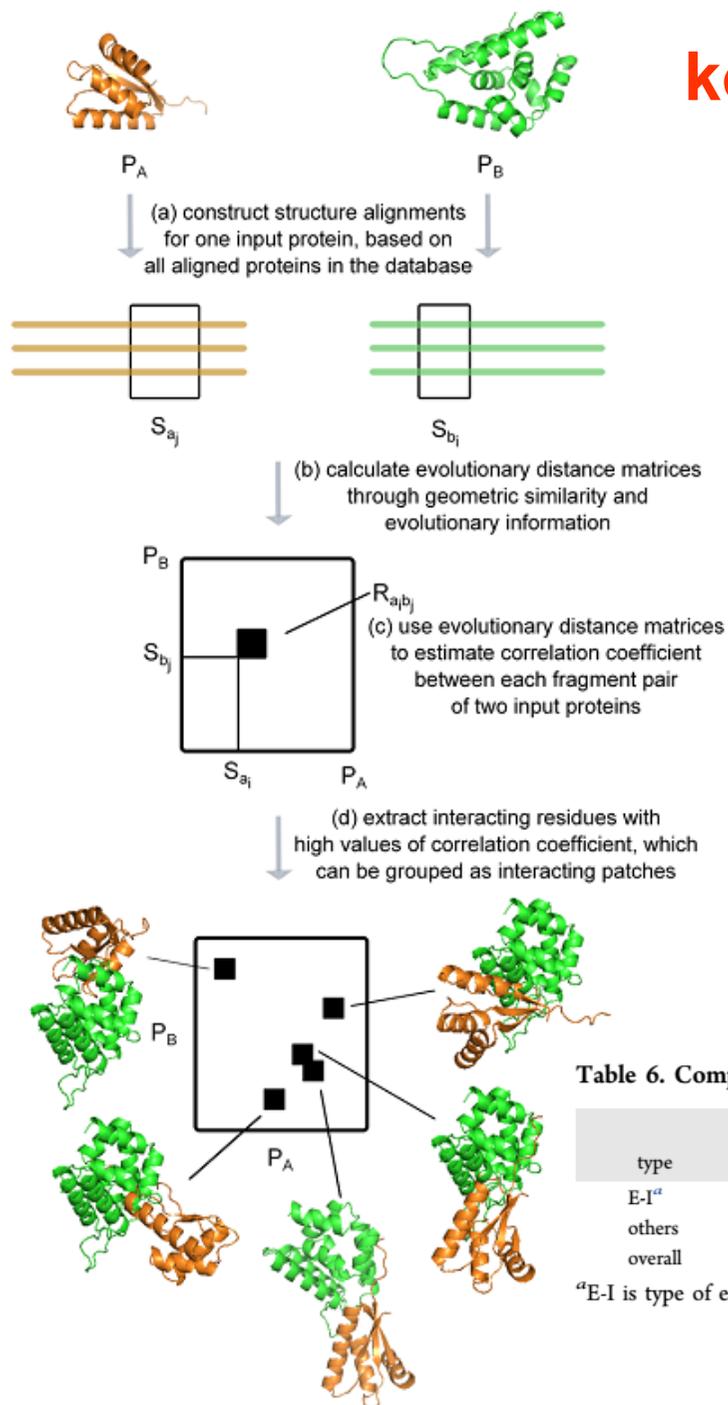
Falls an einer Schnittstelle eine Mutation auftritt, die den Charakter der Aminosäure ändert (z.B. hydrophob/hydrophob in P1/P2 -> polar/polar in P1'/P2'),

sollten an dem anderen Interface korrespondierende Mutationen an den Positionen auftreten, die mit der ersten Aminosäure Kontakte bilden.



Die Suche nach solchen korrelierten Mutationen kann dabei helfen, Bindungskandidaten zu finden.

# korrelierte Mutationen



**Figure 3.** Our method detects correlated residues on SK/RR interaction. Interface residues are indicated in red boxes, and non-interface residues are indicated in black boxes.

**Table 6.** Comparison to metaPPI, meta-PPISP, and PPI-Pred

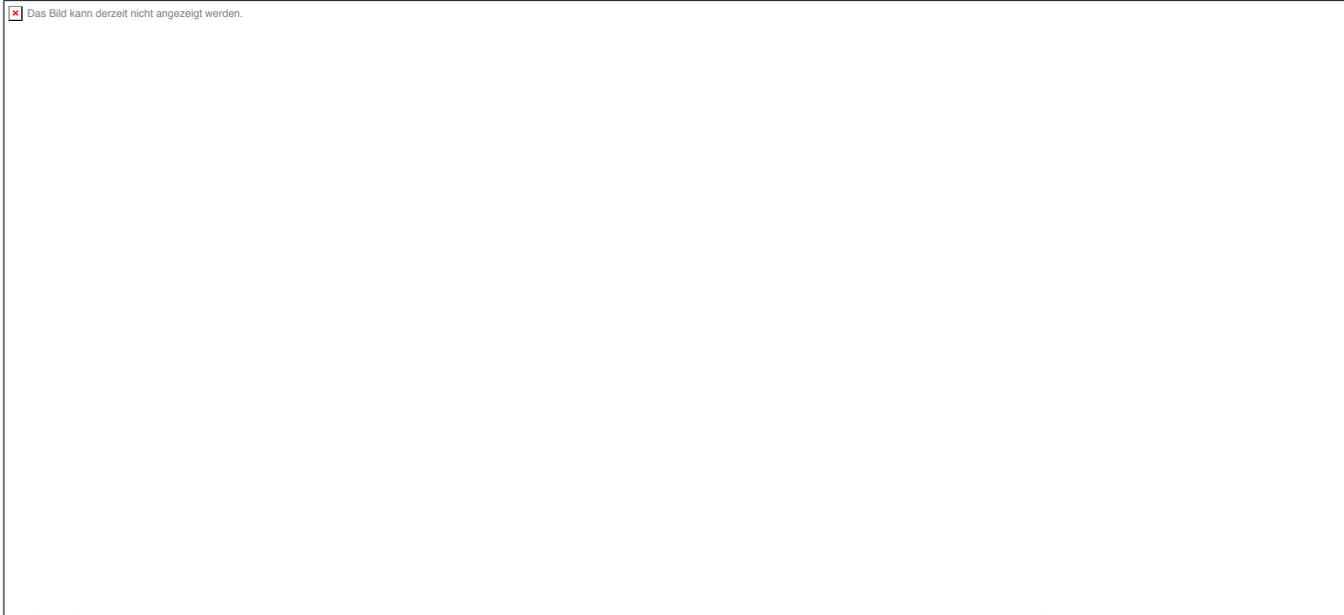
type	our method		metaPPI		meta-PPISP		PPI-Pred	
	Acc	Cov	Acc	Cov	Acc	Cov	Acc	Cov
E-I <sup>a</sup>	73%	65%	61%	37%	56%	55%	46%	47%
others	55%	57%	41%	22%	39%	26%	29%	31%
overall	63%	61%	49%	28%	46%	38%	36%	38%

<sup>a</sup>E-I is type of enzyme-inhibitor.

Guo et al. J. Chem. Inf. Model. 2015, 55, 2042–2049

# Link-Vorhersage basierend auf Netzwerkdaten

Kovács, ... Vidal &  
Barabási  
Nature Commun. 10,  
1240 (2019)



(a) In sozialen Netzwerken impliziert eine große Anzahl an gemeinsamen Freunden eine hohe  $W$ 'keit, dass 2 Leute Freunde werden (rote Verbindung zwischen Knoten X und Y). Dies nennt man **Triadic Closure Principle (TCP)**.

TCP sagt basierend auf einer Knotenähnlichkeit ( $S$ ) Verbindungen ( $P$ ) voraus. Ein Maß für die Ähnlichkeit ist z.B. die Anzahl an gemeinsamen Nachbarn zwischen jedem Knotenpaar ( $A^2$ ).

(b) Übertragen auf PPI-Netzwerke würde man erwarten, dass Proteinpaare mit hoher Jaccard-Ähnlichkeit ebenfalls miteinander interagieren.

# TCP trifft nicht auf PPI-Netzwerke zu

Kovács, ... Vidal &  
Barabási  
Nature Commun. 10,  
1240 (2019)

Jaccard-Koeffizient:  
 $J = |N_X \cap N_Y| / |N_X \cup N_Y|$ ,  
wobei  $N_X$  and  $N_Y$  die Anzahl  
an Interaktionspartnern von  
X und Y sind.

Allerdings beobachten Kovács et al. in einem sehr zuverlässigen PPI-Netzwerk für den Mensch (HI-II-14) quasi das Gegenteil:

Proteinpaare mit hoher Jaccard-Ähnlichkeit haben eine geringere  $W$ 'keit, miteinander zu interagieren.

# PPIs involvieren Bindungsschnittstellen



PPIs benötigen meist komplementäre Schnittstellen. Deshalb teilen sich zwei Proteine, X und Y, mit ähnlichen Schnittstellen oft viele Bindungspartner.

Allerdings garantiert eine gemeinsame Schnittstelle nicht, dass X und Y direkt miteinander interagieren.

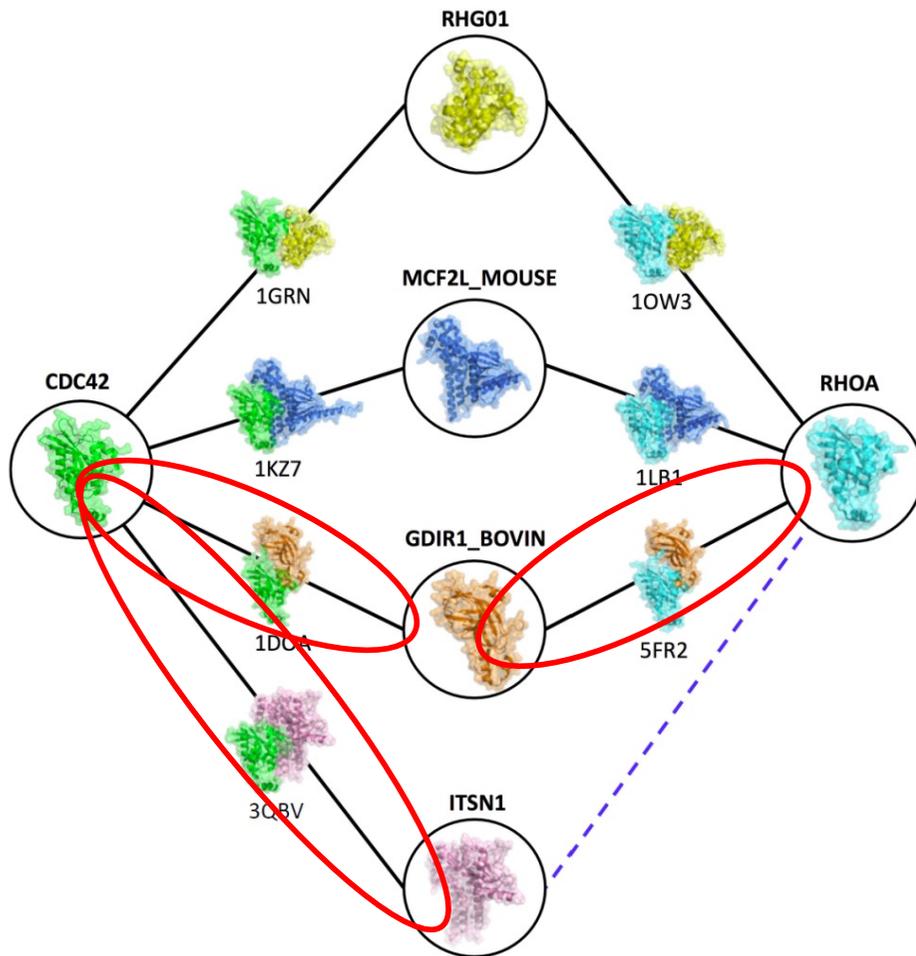
Stattdessen könnte ein weiterer Interaktionspartner von X (Protein D) ebenfalls mit Protein Y interagieren (**blaue Verbindung**).

Solche Kanten können durch Pfade der Länge 3 vorhergesagt werden (L3).

L3 identifiziert ähnliche Knoten zu den bekannten Partnern ( $P = AS$ ), und geht dabei einen Schritt weiter als das Ähnlichkeitsmaß TCP.

Kovács, ... Vidal &  
Barabási  
Nature Commun. 10,  
1240 (2019)

# Strukturelle Veranschaulichung von L3



Kovács, ... Vidal &  
Barabási  
Nature Commun. 10,  
1240 (2019)

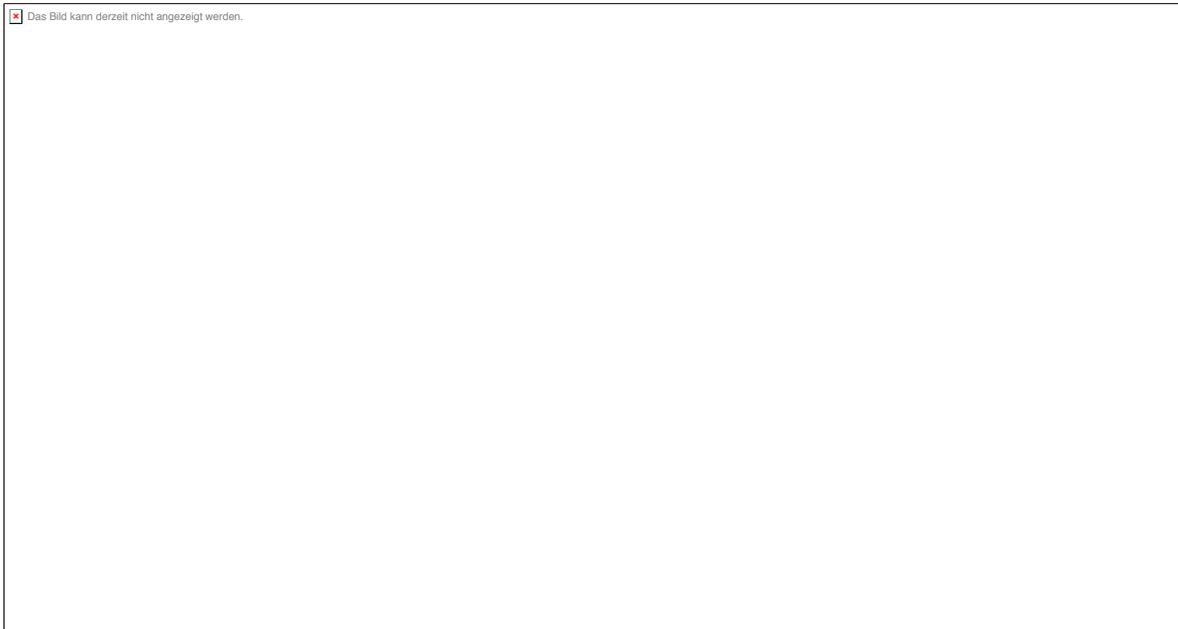
Gezeigt sind PDB-Strukturen für zwei menschliche Proteine, CDC42 und RHOA, die mit manchen Interaktionspartnern durch das gleiche, gemeinsame Interface wechselwirken.

CDC42 und RHOA wechselwirken nicht miteinander. Sie könnten aber weitere gemeinsame Interaktionspartner haben, die an das gemeinsame Interface binden.

Z.B. deutet die **blau** gestrichelte Kante eine mögliche Interaktion zwischen ITSN1 und RHOA an.

Es existieren eine große Anzahl an Pfaden der Länge  $l = 3$  in dem PPI-Netzwerk zwischen ihnen. Hier gezeigt sind 3 Pfade.

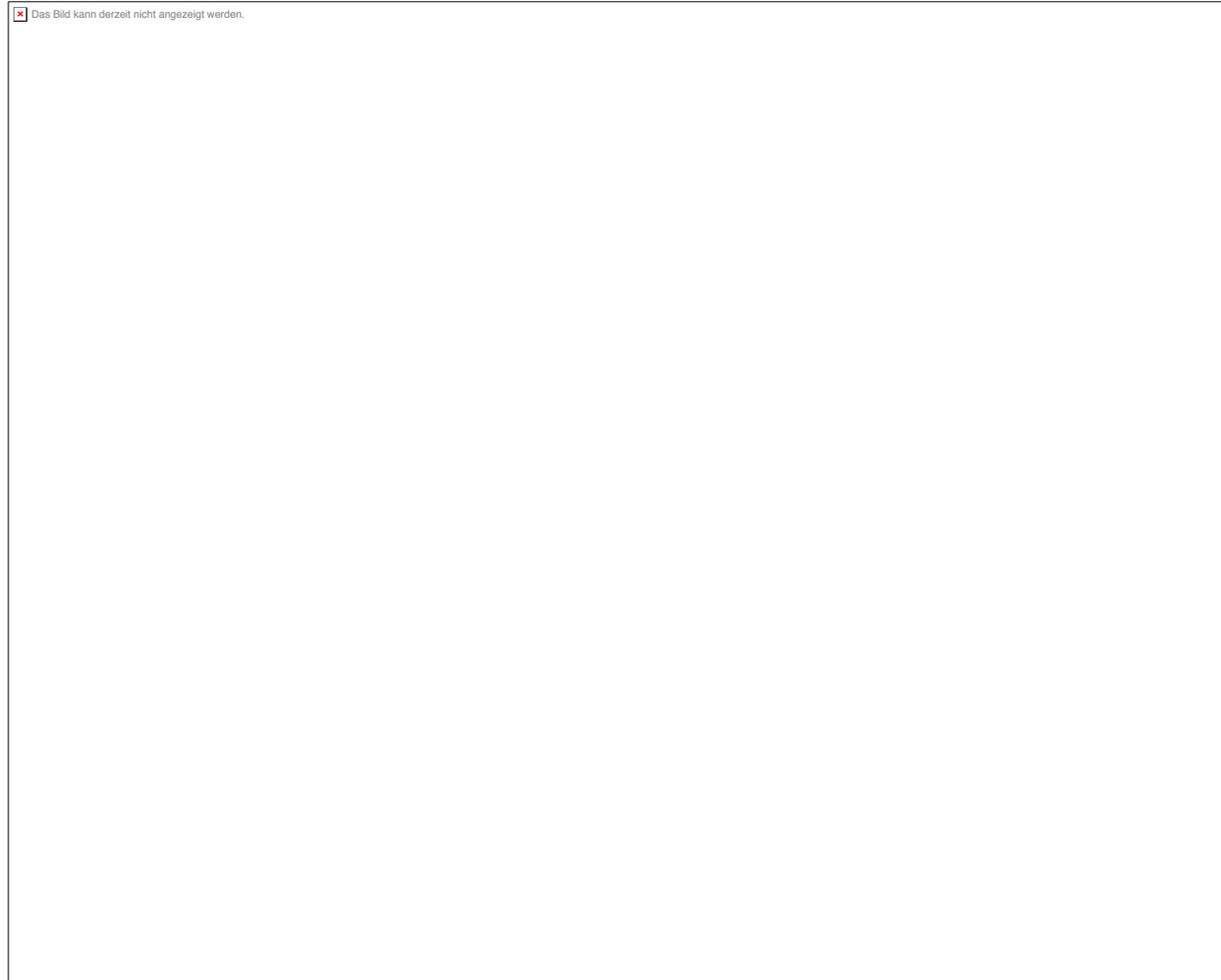
## L3 trifft auf PPI-Netzwerke zu!



**e** Selbst ohne existierende Strukturinformation kann man erwarten, dass 2 Proteine Y und D miteinander interagieren, wenn sie durch mehrere Pfade der Länge  $\ell = 3$  im PPI-Netzwerk verbunden sind. (L3).

**f** Für das Benchmark PPI-Netzwerk HI-II-14 für menschliche Proteine wurde eine positive Korrelation entsprechend **e** beobachtet.

# Durch L3 vorhergesagte Interaktion



FAM161A and PRPF31 sind 2 Proteine, die mit der Krankheit *retinitis pigmentosa* verknüpft werden.

Gezeigt sind alle bekannten Interaktionspartner (grau), zusammen mit den durch L3 vorhergesagten (blau).

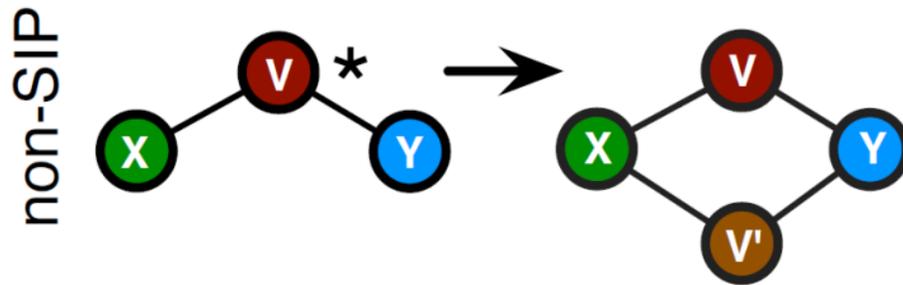
Die stärkste durch L3 vorhergesagte Interaktion verbindet FAM161A mit GOLGA2.

Die beiden Proteine haben keine gemeinsamen Interaktionspartner.

Knotengrösse und –farbe symbolisieren die Anzahl an Interaktionen.

Kovács, ... Vidal &  
Barabási  
Nature Commun. 10,  
1240 (2019)

## Verbindung zur Evolution



**Genduplikation** ist ein Schlüsselmechanismus der Evolution, durch den neue Proteine entstehen können.

Wenn Protein V dupliziert wird (bzw. das dafür kodierende Gen), wird der duplizierte Knoten (V') zumindest anfangs die Interaktionen des Originalproteins V mit X und Y behalten.

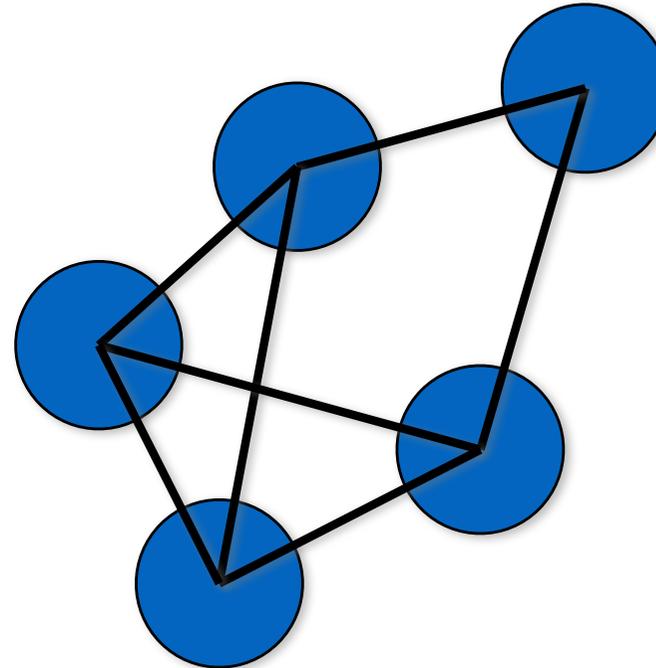
Dies kann den Erfolg von L3 teilweise erklären.

Kovács, ... Vidal &  
Barabási  
Nature Commun. 10,  
1240 (2019)

# Spezifische PP-Interaktionsnetzwerke für bestimmte Bedingungen

## Modell: Brustkrebs (gute Datenlage)

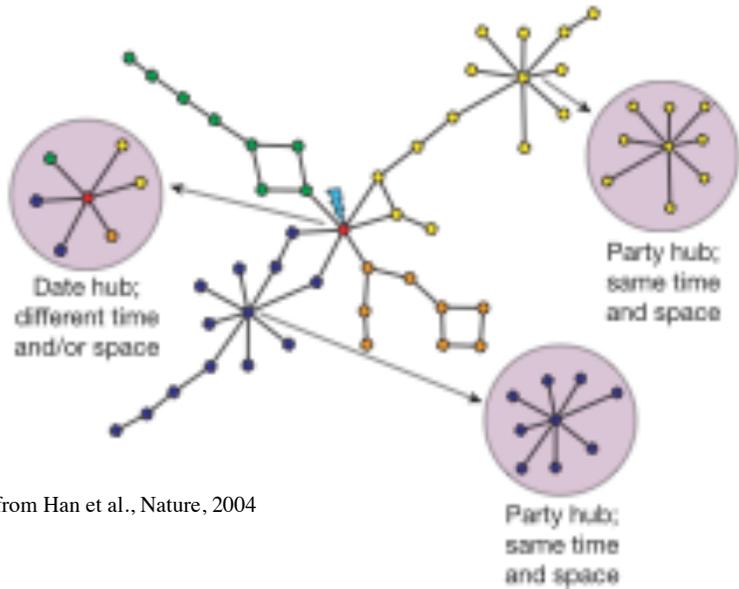
Abb. zeigt das gesamte PPI-  
Netzwerk, z.B. für Mensch  
= Sammlung paarweiser  
Interaktionen aus  
**verschiedenen** Experimenten



Will, Helms, Bioinformatics, 47, 219 (2015)  
doi: 10.1093/bioinformatics/btv620

# Proteininteraktionen können jedoch ...

dynamisch in Zeit and Raum

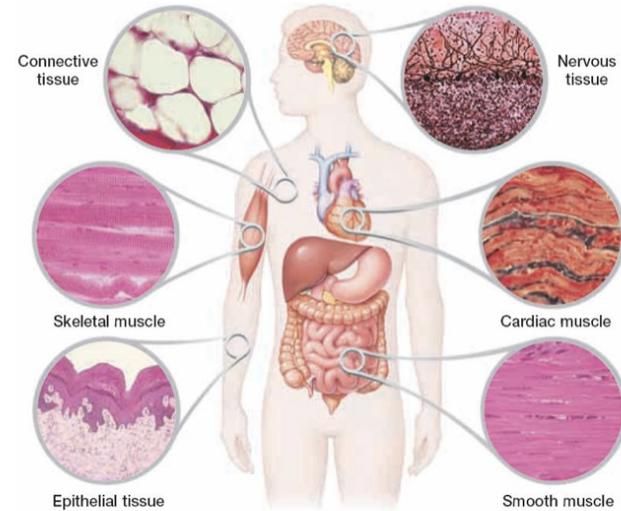


from Han et al., Nature, 2004

Gleiche Farbe = ähnliches Expression spröfil

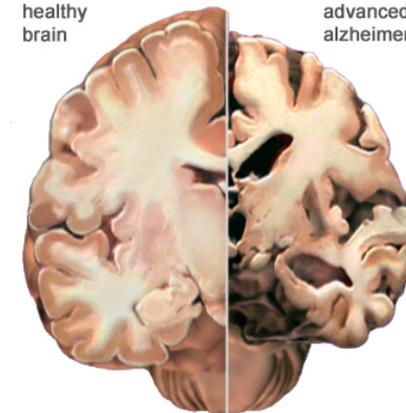
and

Proteinkomposition variiert je nach Bedingung



healthy brain

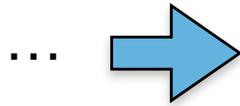
advanced alzheimer's



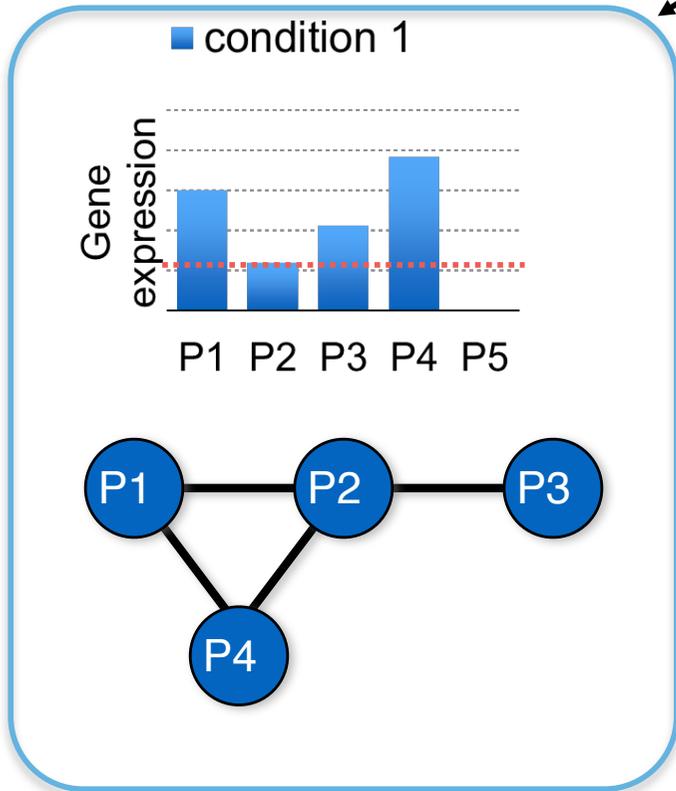
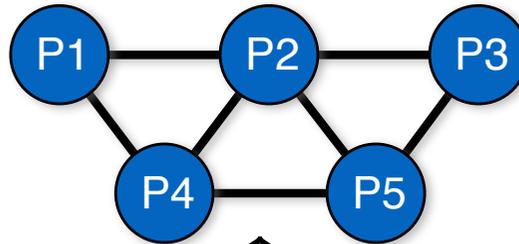
Human tissues from [www.pharmaworld.pk](http://www.pharmaworld.pk)  
Alzheimer from [www.alz.org](http://www.alz.org)

# Gen-basierte Bedingungs-spezifische PPI Netzwerke

Datenbank(en)

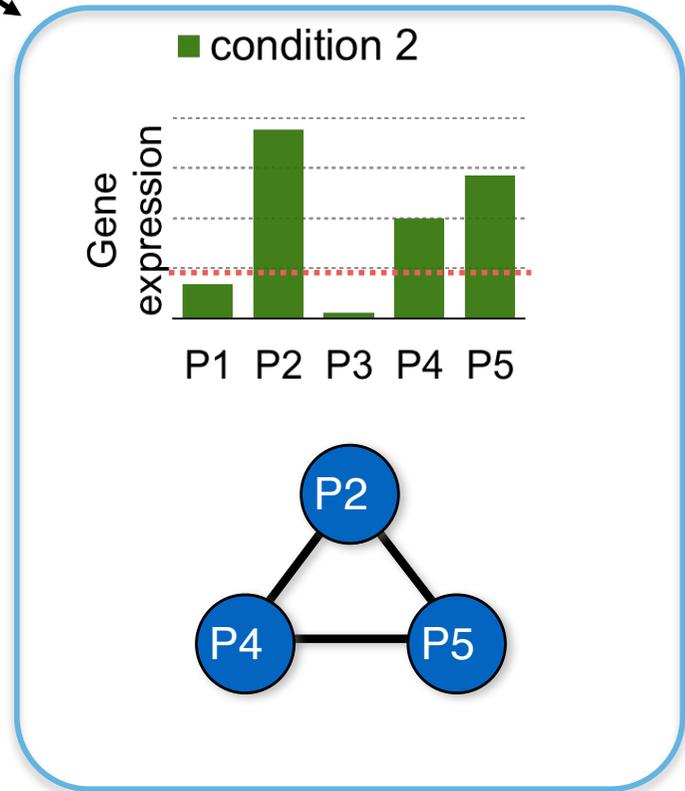


Vollständiges PPI Netzwerk



Idee:  
breche auf Teilmenge  
der exprimierten  
Gene herunter

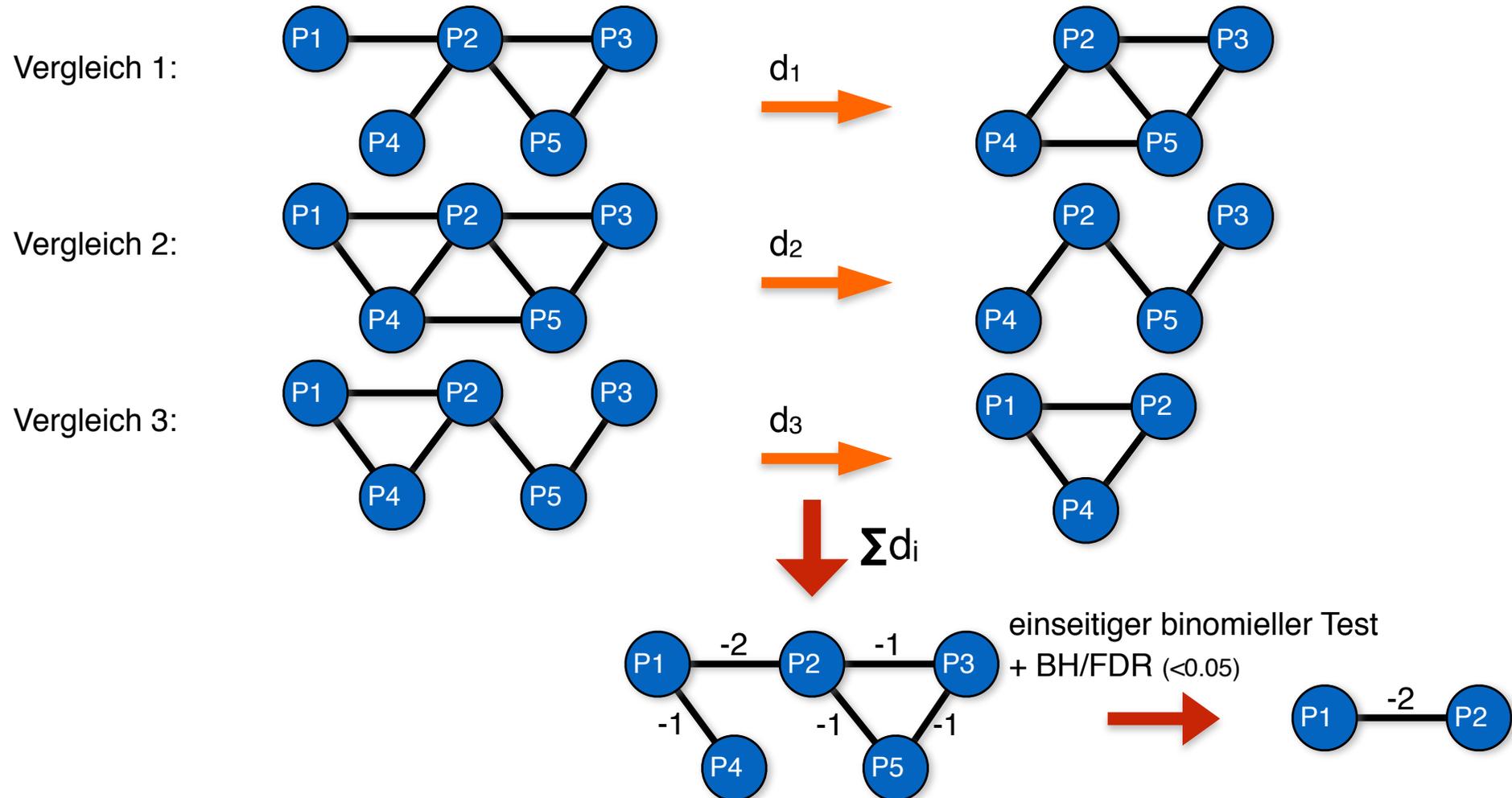
e.g.:  
Bossi and Lehner, Mol. Syst. Bio., 2009  
Lopes et al., Bioinformatics, 2011  
Barshir et al., PLoS CB, 2014



# Differentielle Analyse der PPI (Um)-Verknüpfungen

112 normale Gewebe (TCGA – gleiche Personen))

112 Brustkrebsgewebe (TCGA)



Prüfe, ob eine bestimmte PP-Interaktion in einer signifikant höheren Anzahl an Patienten geändert wird als man zufällig erwartet.

für

## Binomieller Verteilung / Test

Die diskrete Wahrscheinlichkeitsverteilung mit der Wahrscheinlichkeitsfunktion

$$B(k | p, n) = \binom{n}{k} p^k (1 - p)^{n-k} \quad k = 0, 1, \dots, n$$

heißt die *Binomialverteilung* zu den Parametern  $n$  (Anzahl der Versuche) und  $p \in [0, 1]$  (der *Erfolgs-* oder *Trefferwahrscheinlichkeit*).

Ein **Binomialtest** ist ein statistischer Test, bei dem die Testgröße binomialverteilt ist. Er wird verwendet um Hypothesen über Merkmale zu prüfen, die genau zwei Ausprägungen annehmen können.

In Fall der PP-Interaktionen kann eine Interaktion existieren oder nicht.

Man erzeugt also für eine genauso große Anzahl von 112 PP-Netzwerken dieselbe Anzahl von etwa 10.000 zufälligen Verschaltungsänderungen

Dann zählt man ab, wie häufig eine bestimmte Interaktion  $P_i-P_j$  zufällig „rewired“ wird und vergleicht dies mit der in Patienten beobachteten Anzahl.

Daraus erhält man einen p-Wert für diese Interaktion  $P_i-P_j$ .

# Rewiring von PPIs in Brustkrebs vs. gesundem Gewebe

Im Mittel liegen 12.500 – 12.600 Proteine vor.

Die Standardabweichung drückt Unterschiede zwischen einzelnen Patienten aus.

Anhand der bekannten Interaktionsdaten erwartet man zwischen diesen Proteinen etwa 134.000 PP-Interaktionen

-> etwa 10.000 dieser PP-Interaktionen sind in Krebs-Gewebe signifikant anders verschaltet als in gesundem Gewebe.

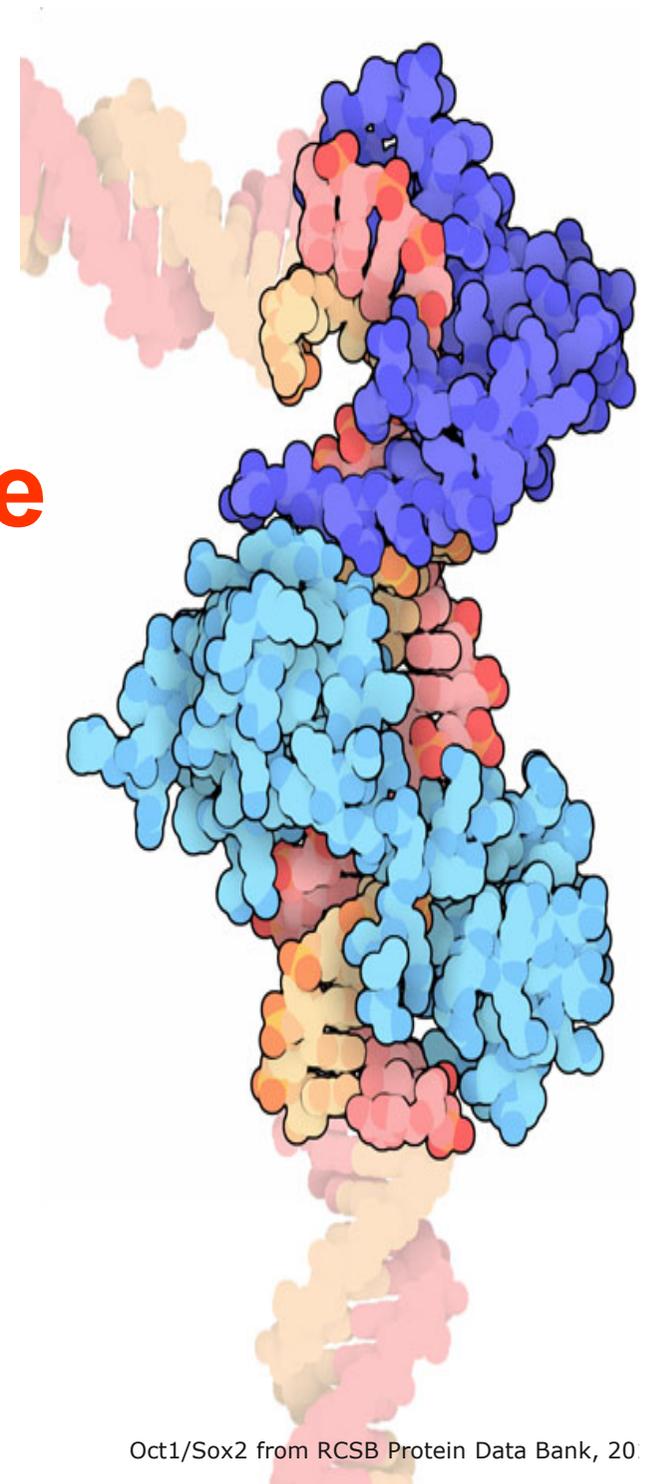
	GENE
avg. number of proteins (normal)	12, 678 ± 223
avg. number of proteins (tumor)	12, 528 ± 206
avg. number of interactions (normal)	134, 348 ± 2, 387
avg. number of interactions (tumor)	133, 128 ± 2, 144
$P_{rew}$	0.067 ± 0.016
significantly rewired interactions	9, 754

# Rewired PPIs sind mit Krebs-Merkmalen assoziiert

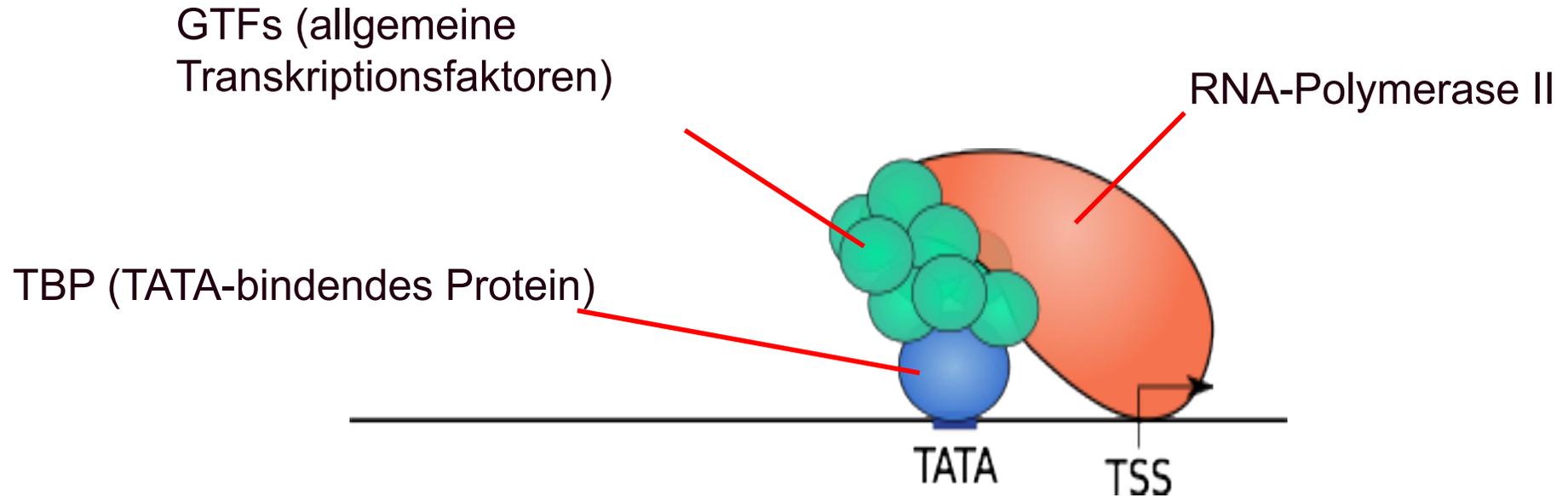
	GENE
rewired interactions	9,754
participation in any hallmark term	7,028
fraction in any hallmark term	0.721
Resisting Cell Death	4,064 (0.417)
Activating Invasion and Metastasis	2,244 (0.230)
Sustaining Proliferative Signaling	3,964 (0.406)
Inducing Angiogenesis	169 (0.017)
Tumor-Promoting Inflammation	516 (0.053)
Genome Instability and Mutation	1,362 (0.140)
Enabling Replicative Immortality	232 (0.024)
Evading Growth Suppressors	3,362 (0.345)
Avoiding Immune Destruction	752 (0.077)
Deregulating Cellular Energetics	821 (0.084)
avg.	1,749 (0.179)

Ein großer Anteil (72%) der anders verschalteten (rewired) Interaktionen betrifft Proteine, die mit den hier aufgeführten, bekannten Merkmalen von Krebs („hallmarks of cancer“) assoziiert sind.

# Transkriptionsfaktor- komplexe in Hefe und ihre Rolle

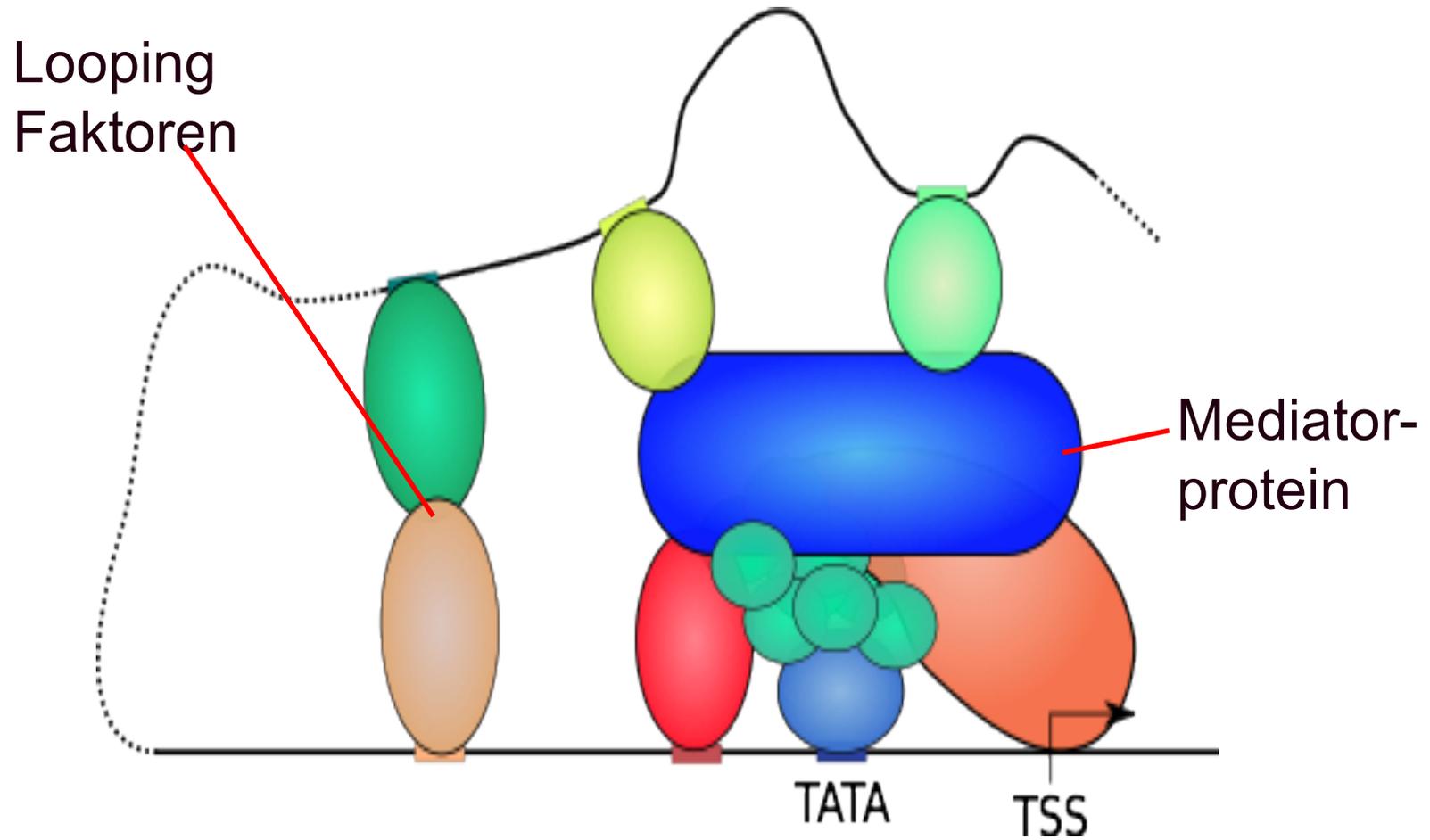


# Transkription: Rolle von TFs

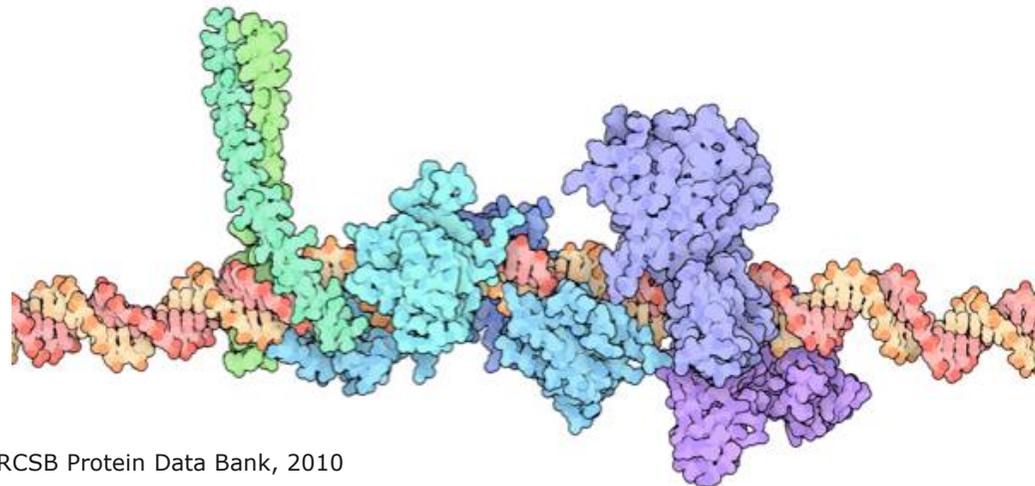
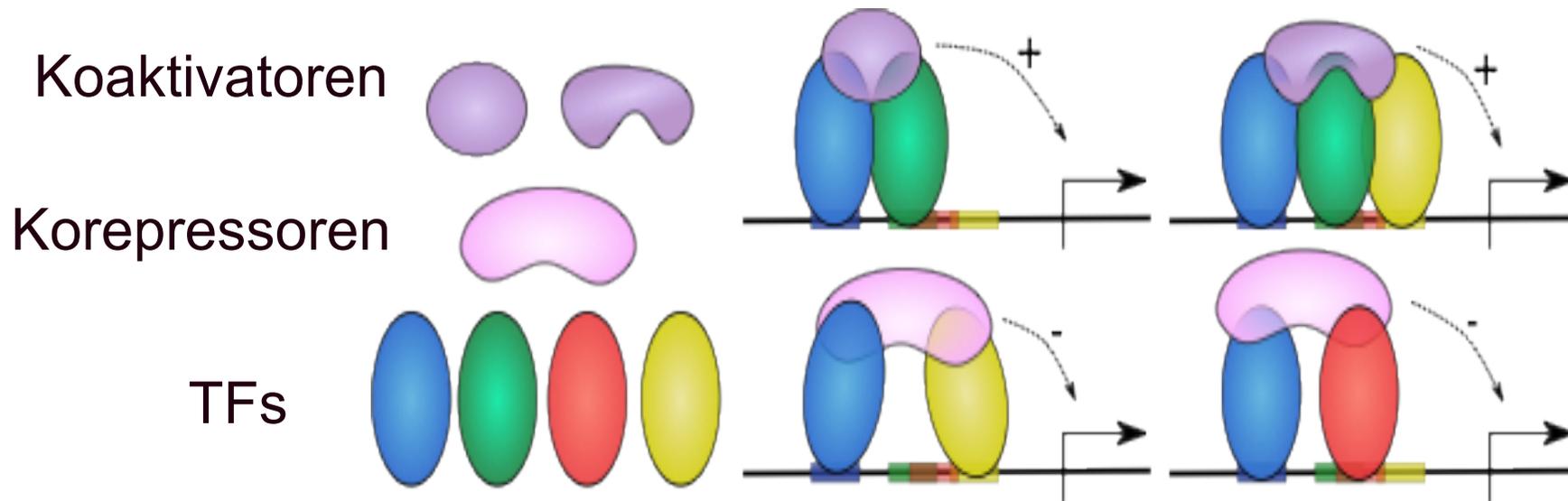


TBP bindet an die DNA, wechselwirkt mit den allgemeinen Transkriptionsfaktoren und rekrutiert RNA-Polymerase II

# Kombinatorische Vielfalt vieler TFs, Bindung weiterer Proteine möglich

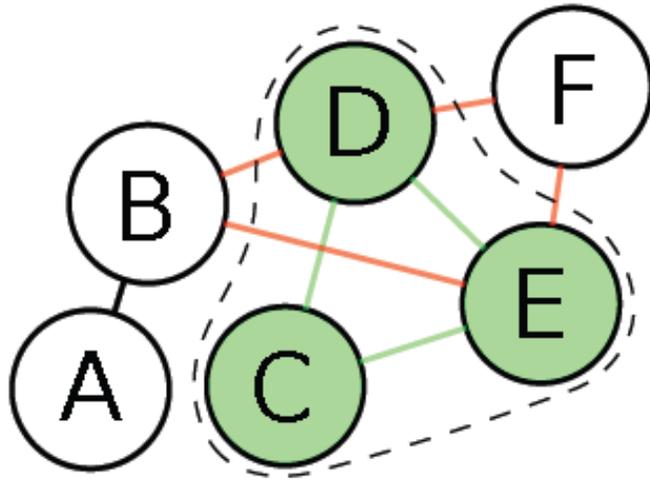


# *cis*-regulatorische Module



IFN-enhanceosome from RCSB Protein Data Bank, 2010

# identifiziere Proteinkomplexe, die TFs beinhalten aus PPI-Netzwerk



Der gestrichelt umrandete  
Komplex C-D-E hat die maximale  
Cohesiveness.

Verwende Idee der Methode ClusterOne:  
Identifiziere Kandidaten für TF-Komplexe im  
Protein-Interaktionsgraph durch  
Maximierung der Cohesiveness  $f(V)$ :

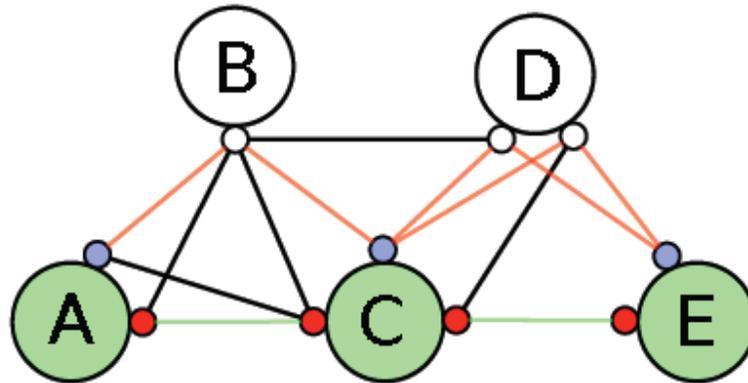
$$f(V) = \frac{w^{in}(V)}{w^{in}(V) + w^{bound}(V)}$$

$w^{in}(V)$ : Summe der internen (gewichteten)  
Kanten, im Beispiel grün

$w^{bound}(V)$ : Summe der externen Kanten, im  
Beispiel orange

# Domänen-Domänen Repräsentation des PPI-Netzwerks

Annahme: jede Domäne kann nur an einer Interaktion beteiligt sein.



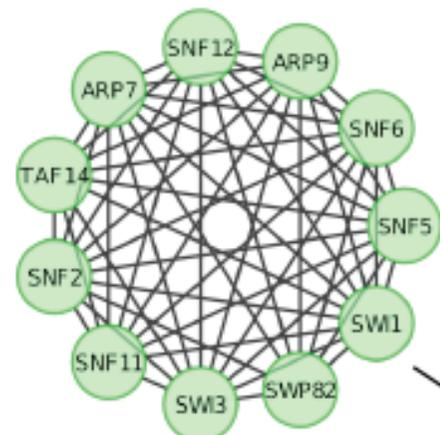
Die grünen Proteine A, C, E bilden aktuellen Komplex.

Die grünen Kanten zwischen ihnen stehen für direkte Kontakte zwischen ihren roten Domänen (kleine Kreise).

B und D sind Kandidaten für einen größeren Komplex.

Ihre weißen Domänen könnten über die orangenen Kanten neue Interaktionen mit den nicht belegten (blauen) Domänen von A, C, E ausbilden.

# Datenquellen



weighted protein-protein interaction data, such as

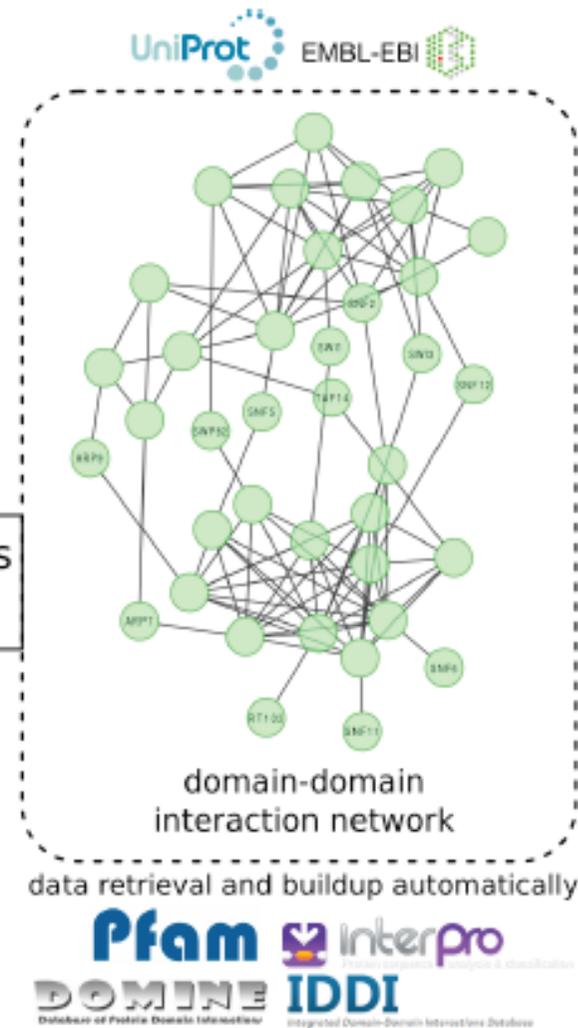


SIF format with UniProt IDs

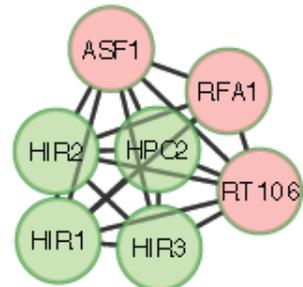
seed proteins,  
threshold for pairs,  
max. depth of search

domain-aware cohesiveness optimization algorithm

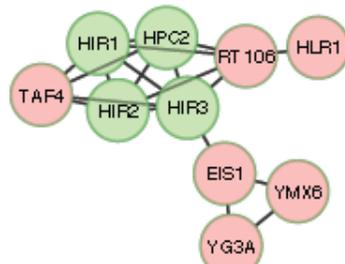
transcription factor complex candidates



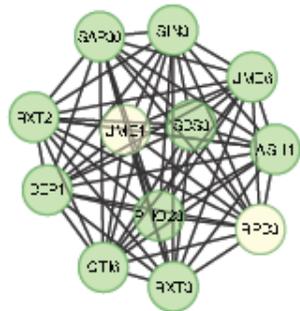
# Beispiele für TF-Komplexe (DACA vs. ClusterONE)



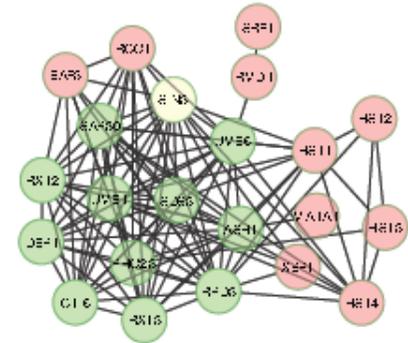
(a) HIR(SGD) / DACA



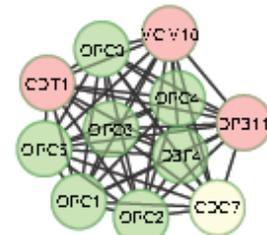
(b) HIR(SGD) / ClusterONE



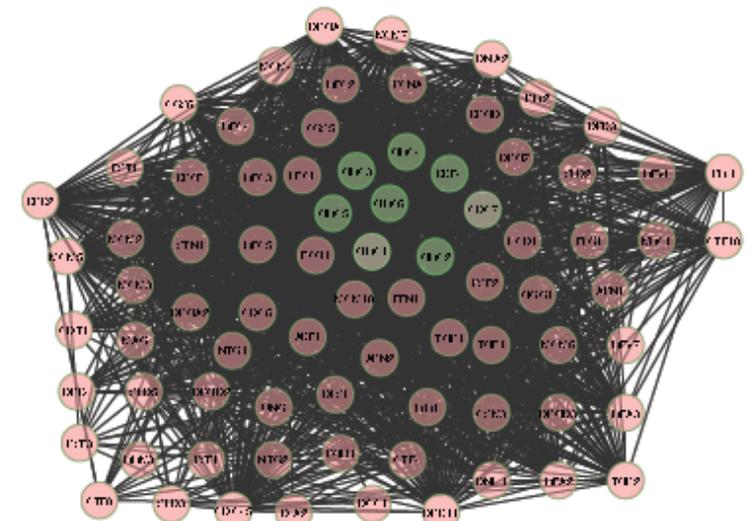
(c) RPD3L(CYC2008) / DACA



(d) RPD3L(CYC2008) / ClusterONE



(e) ORC(MIPS) / DACA

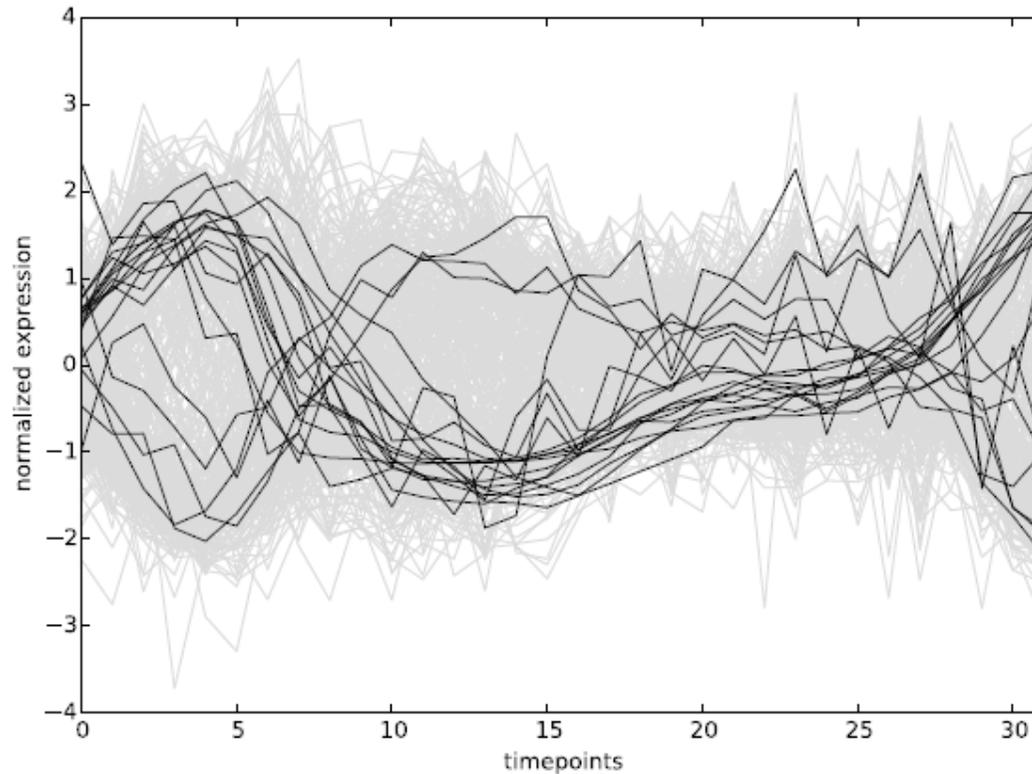


(f) ORC(MIPS) / ClusterONE

Grüne Knoten: Proteine im Referenzkomplex, die korrekt vorhergesagt wurden.

Rote Knoten: weitere Proteine, die als Teil des Komplexes vorhergesagt werden, die aber experimentell nicht bestätigt sind.

# Targetgene von TF Komplexen sind ko-exprimiert!



X-Achse: 32 Zeitpunkte entlang des Zellzyklus von Hefezellen.  
Zellen wurden für Messung synchronisiert.

Grau: Expression der Targetgene von MET4 **oder** MET32 während des Zellzyklus

Schwarz: Expression der Targetgene, die sowohl von MET4 **und** von MET32 reguliert werden.

-> es gibt 2 Gruppen von Targetgenen, die sich sehr ähnlich verhalten.

# Zusammenfassung – PP-Komplexe und Netzwerke

Etwa die Hälfte aller zellulären Proteine beteiligen sich transient oder permanent an Interaktionen mit anderen Proteinen.

Im Mittel interagiert ein Protein mit 6 anderen Proteinen (in Hefe).

Protein-Schnittstellen sind (etwas) stärker **konserviert** als die restliche Protein-Oberfläche (Problem: es kann ja weitere Interaktionen geben ...).

**Korrelierte Mutationen** an Schnittstellen sind starke Indizien für PPIs.

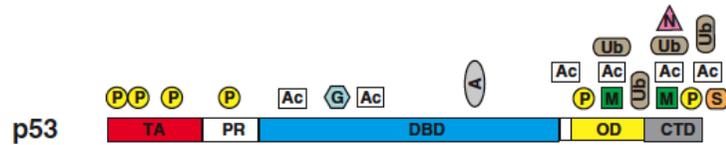
**Hub-Proteine** im Protein-Protein-Interaktionsnetzwerk haben eine höhere Wahrscheinlichkeit **essentiell** zu sein.

Proteinkomplexe, die mehrere Transkriptionsfaktoren erhalten, erhöhen die kombinatorische Vielfalt der Genregulation.

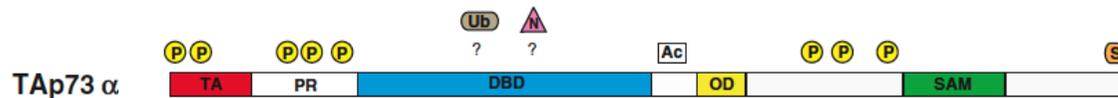
Während der Zelldifferenzierung bzw. Krankheitsentstehung ändern sich eine Vielzahl an Protein-Interaktionen.

# Zusätzliche Folien – nicht benutzt

# Beispiel: das wichtige p53-Protein / post-translationale Modifikationen



p53 („der Wächter der Zelle“) ist ein Transkriptionsfaktor, der ggfs. den Zelltod bzw. die DNA-Reparatur einleitet.



p53 ist bei der Hälfte aller Krebsfälle mutiert, siehe auch Vorlesung V7.



p53 hat im Menschen 2 Verwandte: p63 und p73 mit ähnlichem Domänen-Aufbau.

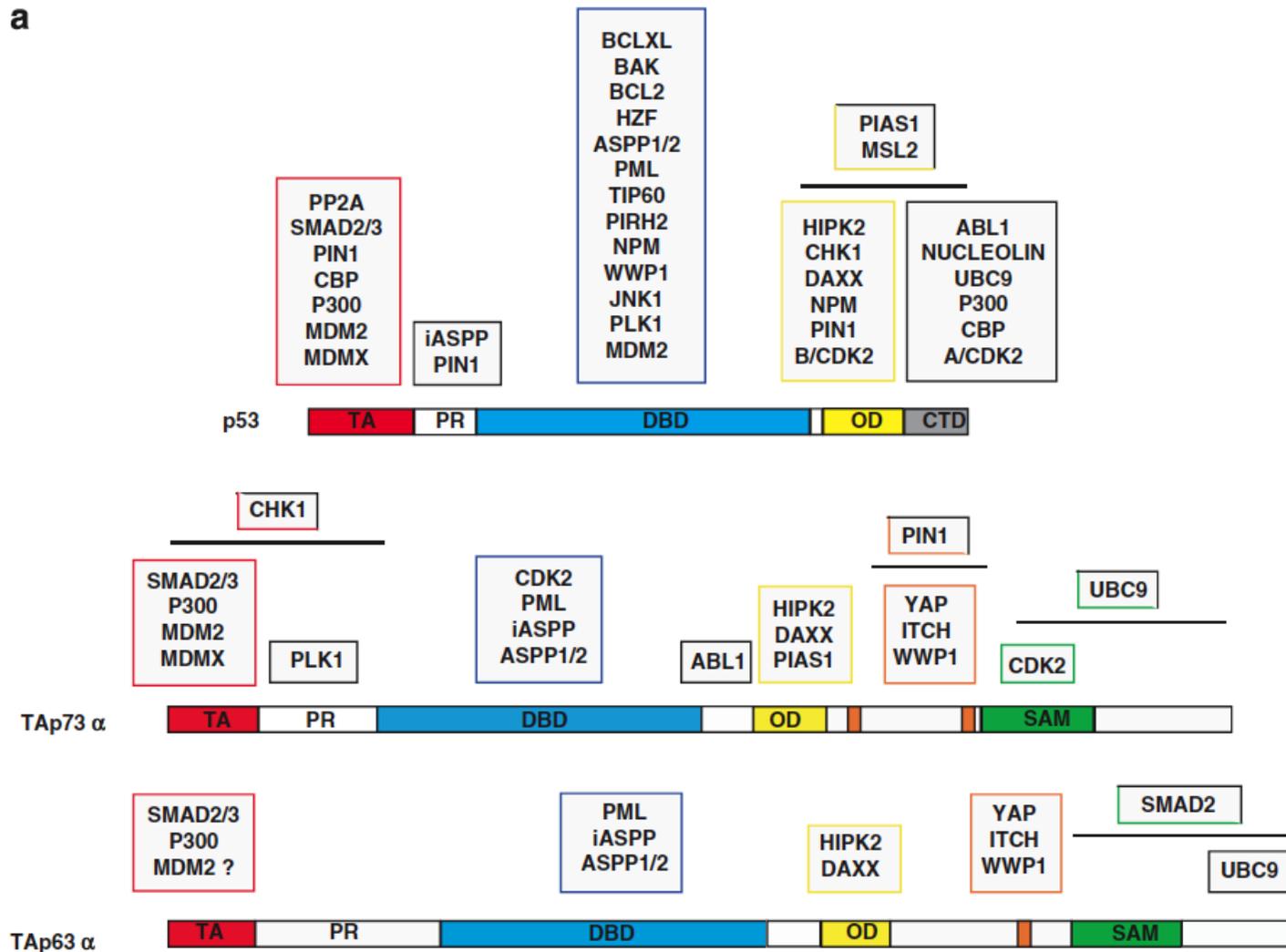
- |                              |                 |
|------------------------------|-----------------|
| Acetylation                  | Phosphorylation |
| Methylation                  | Neddylation     |
| SUMOylation                  | Ubiquitination  |
| O-linked N-Acetylglucosamine |                 |
| poly (ADP-ribose)            |                 |

DBD: DNA binding domain

Collavin et al. Cell Death and Differentiation (2010) 17, 901–911

# Proteininteraktionen von p53, p63 und p73

a



Von p53 ist bekannt, dass es mit 300 Proteinen in der Zelle interagiert.

MDM2 ist ein wichtiger Inhibitor von p53.

Die Interaktionen von p63 und p73 wurden bislang weniger intensiv untersucht.

PP-Interaktionen werden oft durch post-translationale Modifikationen beeinflusst.

Collavin et al. Cell Death and Differentiation (2010) 17, 901–911

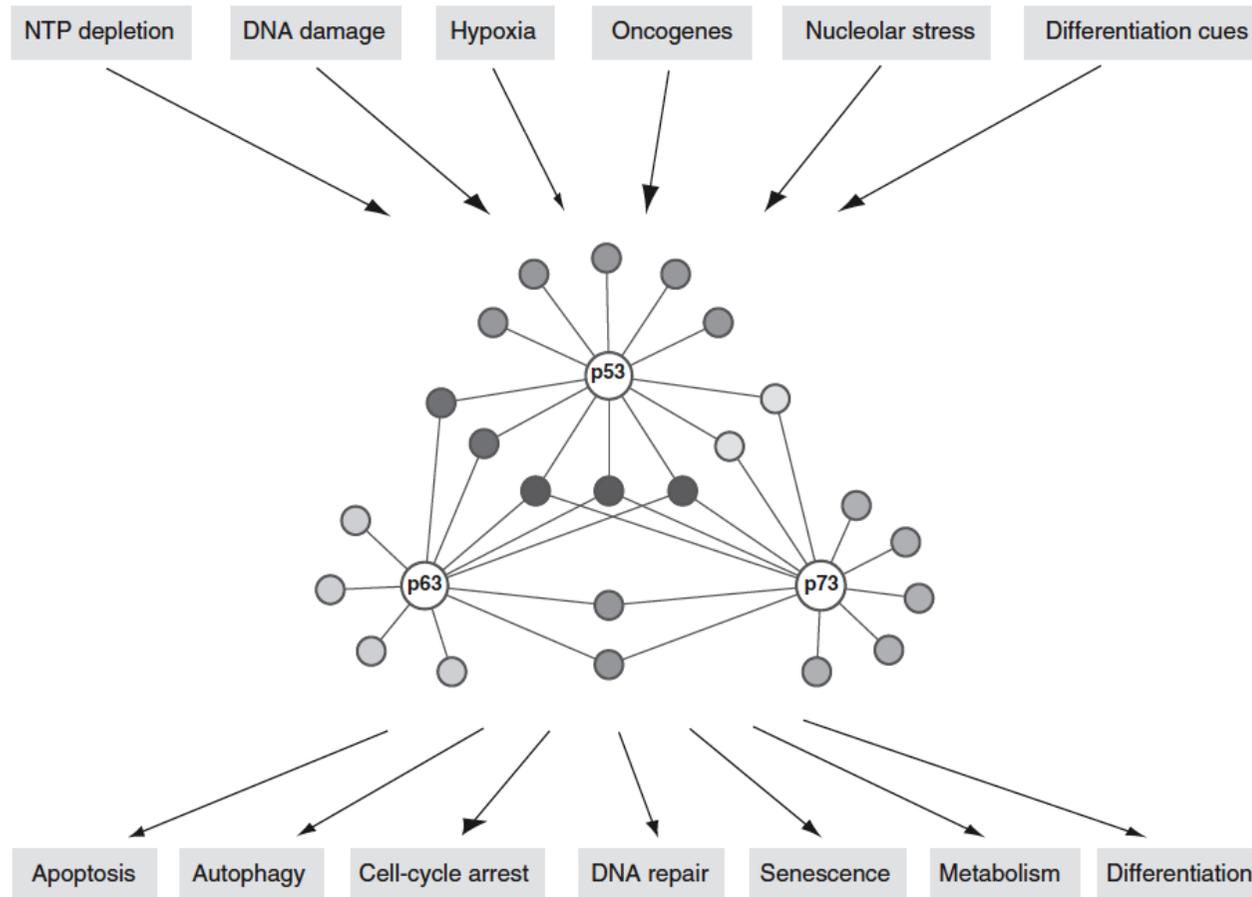
# Enzyme, die Mitglieder p53-Familie modifizieren

**Table 1** A selection of enzymes that apply post-translational modifications on p53-family proteins, or were reported to physically interact with them, independently of modification

	p53	p73	p63	References <sup>a</sup>					
<i>Kinases</i>									
ABL1	+	+	+	30, 67–69					
ATM	+		+	122					
CDK2	+	+	+	122, 123					
CHK1	+	+		124					
CK1	+								
GSK3-beta	+		+	125					
HIPK2	+	+	+	96					
JNK1	+	+		126					
p38	+	+	+	100, 127					
PLK1	+	+		128					
<i>Phosphatases</i>									
PP2A	+		+	125					
Wip1	+								
<i>Acetyltransferases</i>									
P300/CBP	+	+	+	77, 78, 100					
PCAF	+	+		77					
TIP60	+	+		129					
<i>De-acetylases</i>									
HDAC1/2	+	+		130					
Sirt1	+	+		83					
<i>Ubiquitin ligases</i>									
MDM2	+	+	+ /-	44, 49, 50					
MDMX	+	+	-	42, 43, 49					
COP1	+								∞
PIRH2	+								
Synoviolin	+								
ARF-BP1	+								
CHIP	+								
WWP1	+	+	+	63, 64					
ITCH	-	+	+	65, 66					∞
E4F1	+								
TRIM24	+								
FBXO45	-	+		62					
<i>Deubiquitinases</i>									
HAUSP	+								
USP10	+			131					
<i>Methyltransferases</i>									
Smyd2	+								
SET7/8	+								
PRMT5	+								
<i>Demethylases</i>									
LSD1	+								
<i>SUMO ligases</i>									
Ubc9	+	+	+	132, 133					
PIAS1	+	+		134					
TOPORS	+								
<i>Others</i>									
O-GlcN-Ac transferase	+								
PARP-1	+								
PIN1	+	+		73					

Collavin et al. Cell Death and Differentiation (2010) 17, 901–911

# Proteininteraktionsnetzwerke



Die Mitglieder p53-Familie reagieren auf viele zellulären Signale und leiten entsprechende Antworten ein.

Collavin et al. Cell Death and Differentiation (2010) 17, 901–911

# Funktionelle Rolle von TF Komplexen

$P_{dECS} < 0.05$  bedeutet, dass Ko-Expression der Zielgene nicht zufällig ist; vergleiche mit Datenpunkten, bei denen die Zeitpunkte zufällig durcheinander gewürfelt wurden.

TFs	$P_{dECS}$	targets	reg. influence	GO process enrichment ( $P < 0.05$ ) in targets
MET4/MET32	0.0010	19	+	methionine metabolic process
TBP/HAP5	0.0335	47	+	/
GLN3/DAL80	0.0009	28	/	allantoin catabolic process
DIG1/STE12/SWI6	0.0369	15	/	fungal-type cell wall organization
FHL1/RAP1	0.0001	116	+	rRNA transport
RPH1/GIS1	0.0001	100	-	hexose catabolic process
CBF1/MET32	0.0002	33	o	sulfate assimilation
DIG1/STE12	0.0003	34	-	response to pheromone
GCN4/RAP1	0.033	62	+	/
MSN4/MSN2	0.0021	105	+	oligosaccharide biosynthetic process
DAL80/GZF3	0.0044	20	-	purine nucleobase metabolic process
SWI6/SWI4	0.0039	53	+	regulation of cyclin-dependent protein serine/threonine kinase activity
STB1/SWI6	0.0275	47	+	/
TBP/SWI6	0.0159	14	+	/
GLN3/GZF3	0.0120	31	/	allantoin catabolic process
MBP1/SWI6/SWI4	0.0307	18	+	regulation of cyclin-dependent protein serine/threonine kinase activity
MBP1/SWI6	0.0124	25	/	cell cycle process

Jeder Komplex enthält 2-3 TFs (und andere Proteine)

Anzahl an Zielgenen mit Bindemotiven für 2-3 TFs.

Wie bestimmt man diese unter den Zielgenen angereicherten GO-Terme?