

Integrated Genomic and Proteomic Analyses of a Systematically Perturbed Metabolic Network

Trey Ideker, Vesteinn Thorrson, Jeffrey A. Ranish, Rowan Christmas,
Jeremy Buhler, Jimmy K. Eng, Roger Bumgarner, David R. Goodlett,
Ruedi Aebersold, Leroy Hood[1]

Nature 2001

Chris Bielow

Bioinformatik
Freie Universität Berlin



Zusammenfassung

Die Autoren stellen einen neuen Ansatz zur Integration von mRNA- und Protein-expressionsdaten am Beispiel des Galaktoseweges von *Saccharomyces cerevisiae* vor, der es erlaubt bekannte intrazelluläre Stoffwechselwege und Interaktionsnetzwerke zu erweitern und zu verfeinern. Ziel ist es ein besseres Modell des Systems zu entwerfen und die eingeführten Verbesserungen experimentell zu verifizieren. Die Autoren benutzen Microarrayexperimente und Tandem MS/MS und können durch den Vergleich der beiden ca. 15 posttranskriptional modifizierte Proteine identifizieren. Es stellt sich weiterhin heraus, dass 997 Gene in *S. cerevisiae* auf 20 systematische Störungen des GAL-Systems mit einer signifikant veränderten mRNA-Expression reagieren und es somit eine starke Interaktion zwischen dem Galaktoseweg und anderen metabolischen Stoffwechselwegen zu geben scheint. Die Autoren verwenden deshalb bekannte Protein-DNA- und Protein-Protein-Interaktionen um daraus ein großes Interaktionsnetzwerk zu erstellen. Dieses Netzwerk wird um neue hypothetische Interaktionen erweitert und teilweise durch Experimente verifiziert.

1 Einleitung

Ein Weg um die Funktion von Genen und Proteinen besser zu verstehen ist die Erstellung von Interaktionsnetzwerken. Dafür werden oft Yeast-Two-Hybrid, Microarrays, RT-PCR und viele andere Methoden eingesetzt. Die Ergebnisse von Gygi et al.[2] zeigen jedoch, dass die Untersuchung von mRNA-Expressionmustern z.B. durch Microarrays oder Proteinexpressionsmustern z.B. durch MS/MS für sich allein, nicht ausreicht um ein intrazelluläres Netzwerk zu analysieren, da es keine signifikante Korrelation zwischen mRNA- und Proteinexpression zu geben scheint. Gründe für diese fehlende Korrelation können u.a. posttranskriptionale Kontrollmechanismen, Halbwertszeiten von mRNAs und Proteinen oder intrazelluläre Lokalisierung und Bindung der Proteine sein.

Die Autoren beschreiben eine Herangehensweise für die Aufklärung von intrazellulären Netzwerken am Beispiel von *Saccharomyces cerevisiae*, die es erlaubt ein (bekanntes) Netzwerk zu verifizieren und zu erweitern. Das Genom der Bäckerhefe ist vollständig sequenziert. Das bedeutet zwar noch nicht, dass alle Gene vollständig charakterisiert sind¹, lässt aber erste Microarrayanalysen zu. Für das tiefere Verständnis von Netzwerkstrukturen und Interaktionen werden jedoch mehr Informationen benötigt, als man durch alleinige Messung der mRNA-Expression extrahieren kann. Ein möglicher Ansatz ist eine Integration von mRNA- und Proteinexpressionsmessungen.

Von den Autoren wurde dafür der gut untersuchte Galaktoseweg der Hefe gewählt. Er existiert in den meisten Hefestämmen (u.a. auch in der Bäckerhefe) und dient als ein möglicher Lieferant von Glukose für die Glykolyse. Das GAL-System ist ein klassisches Modell eines eukaryotischen Genschalters, der von Galaktose aktiviert und von Glukose gehemmt werden kann. Im Vergleich zu einer vorhergehenden Arbeit von Gygi et al. führen die Autoren vergleichende Analysen zwischen dem Wildtyp und genetisch veränderten Bäckerhefekulturen durch und erhalten damit Hinweise auf die Funktion einzelner Gene/Proteine.

Obwohl der Galaktoseweg sehr ausführlich untersucht wurde ist das bisherige Modell teilweise unvollständig. Abbildung 1 zeigt den Kern der bisher bekannten Informationen. Die Rolle von GAL6 ist hier zum Beispiel noch nicht vollständig geklärt, da es ursprünglich in einem Drug-resistance Pathway entdeckt wurde. Ostergaard et al. vermuten, dass es an der Degradation von Gentranskripten der GAL-Gene beteiligt sein könnte[3]. Am Ende des Galaktoseweges steht die Bereitstellung von Glukose-6-Phosphat für die Glykolyse. Galaktose wird über einen Galaktosetransporter (GAL2) oder über einen unspezifischen Hexosetransporter (HXT) in die Zelle transportiert und dann von einer Reihe von Enzymen (GAL1, GAL7, GAL10, GAL5) zu Glukose-6-Phosphat umgewandelt. Die Expression dieser Gene wird über einen Transkriptionsfaktor (GAL4) gesteuert. Ist er aktiv, so werden die Gene exprimiert. GAL4 wird seinerseits durch einen Faktor (GAL80) inhibiert, der zwischen Zellkern und Cytoplasma hin und her wandern kann. In Anwesenheit von Galaktose wird GAL80 durch einen weiteren Faktor (GAL3), der ausschließlich im Cytoplasma vorkommt, gebunden. GAL80 kann in diesem Zustand nicht mehr in den Zellkern wandern um GAL4 zu inhibieren. Die Anwesenheit von Galaktose in der Zelle führt also zur Aktivierung der GAL-Gene, sofern keine inhibierenden Zucker wie z.B. Glukose vorhanden sind. Andere Energiequellen die der Hefe zu Verfügung stehen sind u.a. Ethanol und Raffinose.

¹derzeit sind ca. 66,24%(4375) der proteinkodierenden Gene verifiziert (Stand: 07.03.2006); alle anderen Gene sind vorhergesagt oder nicht charakterisiert; URL:<http://www.yeastgenome.org/cache/genomeSnapshot.html>; die Menge der bestätigten Gene zum Zeitpunkt der Publikation des hier behandelten Papers ist höchstwahrscheinlich kleiner, enthielt jedoch alle Gene des Galaktoseweges

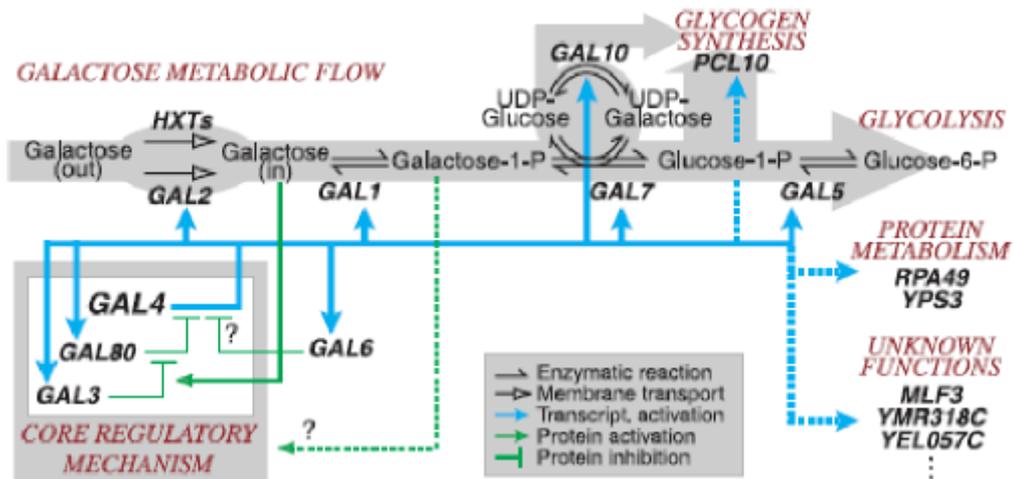


Abbildung 1: Bisher bekanntes Modell des Galaktoseweges einschließlich der von den Autoren vorgeschlagenen Verfeinerungen - dargestellt als gepunktete Linien (Abbildung nach [1]).

2 Methoden

Der von den Autoren verwendete Ansatz sieht die vier folgenden Stufen vor:

- (i) Beschreibung des bisher bekannten Netzwerkes aus Genen, Proteinen und weiteren Molekülen und ihren Interaktionen
- (ii) jede Komponente des Netzwerkes einzeln stören (z.B. durch Mutationen oder veränderte Umweltbedingungen) und Veränderungen auf mRNA und Proteinebene global messen
- (iii) neue Informationen in das Netzwerk integrieren
- (iv) Hypothese über bisher unbekannte Beobachtungen formulieren und mit geeigneten Experimenten verifizieren

Im ersten Schritt (i) betrachten die Autoren das bisher gültige Modell des GAL-Systems, das in der Einleitung beschrieben wurde.

Schritt (ii) sieht die systematische Störung der bisher bekannten Netzwerkkomponenten vor. Um Informationen über das Netzwerk zu erweitern wurden verschiedene Experimente durchgeführt. Es wurden jeweils der Wildtyp und neun mutierte (aber funktionale) Hefestämme in galaktosearmem (-gal) bzw. galaktosereichem (+gal) Medium auf ihre Gen- und Proteinexpression untersucht. Insgesamt ergeben sich dadurch 20 Experimente. In allen Fällen konnten die Expressionslevel aller GAL-Gene und ihrer Produkte gemessen werden. Den mutierten Stämmen fehlt jeweils eines der neun GAL-Gene, durch das sie jeweils ihren Namen erhalten: gal1 Δ , gal2 Δ , gal3 Δ , gal4 Δ , gal5 Δ , gal6 Δ , gal7 Δ , gal10 Δ , gal80 Δ .

Für die Genexpressionsanalyse wurden Microarrays mit je 6200 Genen verwendet, wobei parallel die Expression eines Genes in einer Mutante gegen die Expression eines Genes im Wildtyp unter gleichen äußeren Bedingungen gemessen wurde. Zusätzlich wurde als Negativkontrolle der Wildtyp im +gal Medium gegen sich selber und gegen den Wildtyp im -gal

Medium gemessen. Um eine robuste Messung zu gewährleisten wurden die Experimente jeweils vier mal wiederholt. Daraus lässt sich eine sogenannte mRNA-Störungsmatrix (siehe Abb. 2) erstellen.

Für die Proteinexpressionsmessung wurden ICAT-markierte Zellproben durch eine mehrdimensionale Chromatographie in Verbindung mit einer Tandem MS/MS getrennt und analysiert². Im Gegensatz zur mRNA-Messung wurden hier nur die Daten des Wildtyps in den zwei verschiedenen Kulturmedien erhoben.

3 Ergebnisse & Diskussion

Zum Zeitpunkt der Veröffentlichung war noch nicht hinreichend bekannt, dass Hefezellen nachdem sie veränderten externen Bedingungen ausgesetzt werden, langfristig zu einem Steady State unabhängig vom Nährmedium übergehen. Das war bis dato nur von Hefekulturen unter Stressbedingungen[4] bekannt. Die Autoren haben nicht angegeben wie lange die Hefezellen dem jeweiligen Medium ausgesetzt waren bzw. nach wie vielen Generationen die Zellen analysiert wurden und insbesondere ob das Zellmaterial für die Microarrayexperimente und Proteinexpressionsmessungen zum gleichen Zeitpunkt entnommen wurde. Braun et al.[5] haben im Jahr 2004 gezeigt, dass Hefekulturen nach einer gewissen Zeit ihre Genexpression unabhängig von externen Einflüssen auf einen Steady State einstellen. Diese Entdeckung könnte die Aussagekraft der gewonnenen Daten verringern.

Betrachtet man die aus den Microarraydaten resultierende mRNA-Störungsmatrix (siehe Abb. 2), dann sieht man dass die meisten der antizipierten Effekte der Genknockouts (nach dem bisherigen Modell) auch eingetreten sind. Beispielsweise ist die Expression der meisten GAL Gene vom -gal zu +gal Medium in Wildtyp angestiegen. Das würde man auch erwarten, da das GAL-System durch Galaktose aktiviert wird. Die Konzentration von GAL4 sollte sich dabei wie vorhergesagt nicht ändern, da es konstitutiv exprimiert wird. Weiterhin erwartet man, dass beim Fehlen des Inhibitors GAL3, GAL80 ungehindert und unabhängig vom Galaktosegehalt an GAL4 binden kann und somit die Expression der von GAL4 abhängigen GAL-Gene verhindert wird. Der gleiche Fall tritt ein, wenn das GAL4-Gen direkt ausgeschaltet wird. Beispiele die nicht dem Modell folgen werden am Ende besprochen.

Vergleicht man die globale Gen- und Proteinexpression des Wildtyps in -gal gegen +gal Bedingungen im Medium so findet man lediglich eine mäßige Korrelation ($r = 0.61$), wobei sich bestimmte Gruppen von Proteinen anders verhalten als andere Gruppen. Zum Beispiel zeigen GAL-Gene meist eine positive Korrelation, wobei jedoch ein Anstieg im mRNA-Level nur einen kleinen Anstieg auf Proteinebene nach sich zieht. Die Autoren haben in einer darauf folgenden Arbeit[6] ebenfalls solche Experimente durchgeführt und dort unter etwas veränderten äußeren Bedingungen eine ähnliche Korrelation festgestellt.

Besonders Gene für ribosomale Proteine scheinen einer strikten posttranskriptionalen Kontrolle zu unterliegen, da selbst bei erhöhtem drei- bis fünffachem mRNA Niveau im +gal Medium die Proteinkonzentration stabil zu sein scheint (siehe Abb. 3). Andererseits fanden die Autoren auch 15 Proteine deren mRNA-Level unter verschiedenen Bedingungen konstant blieben, die jedoch eine signifikant andere Proteinkonzentration aufwiesen. Das deutet ebenfalls auf eine posttranskriptionale Kontrolle hin und belegt wie wichtig eine integrierte Analyse ist.

²eine gute Beschreibung der Methode findet sich unter <http://www.lcpackings.com/applications/UltiMate/Proteomics9.htm>

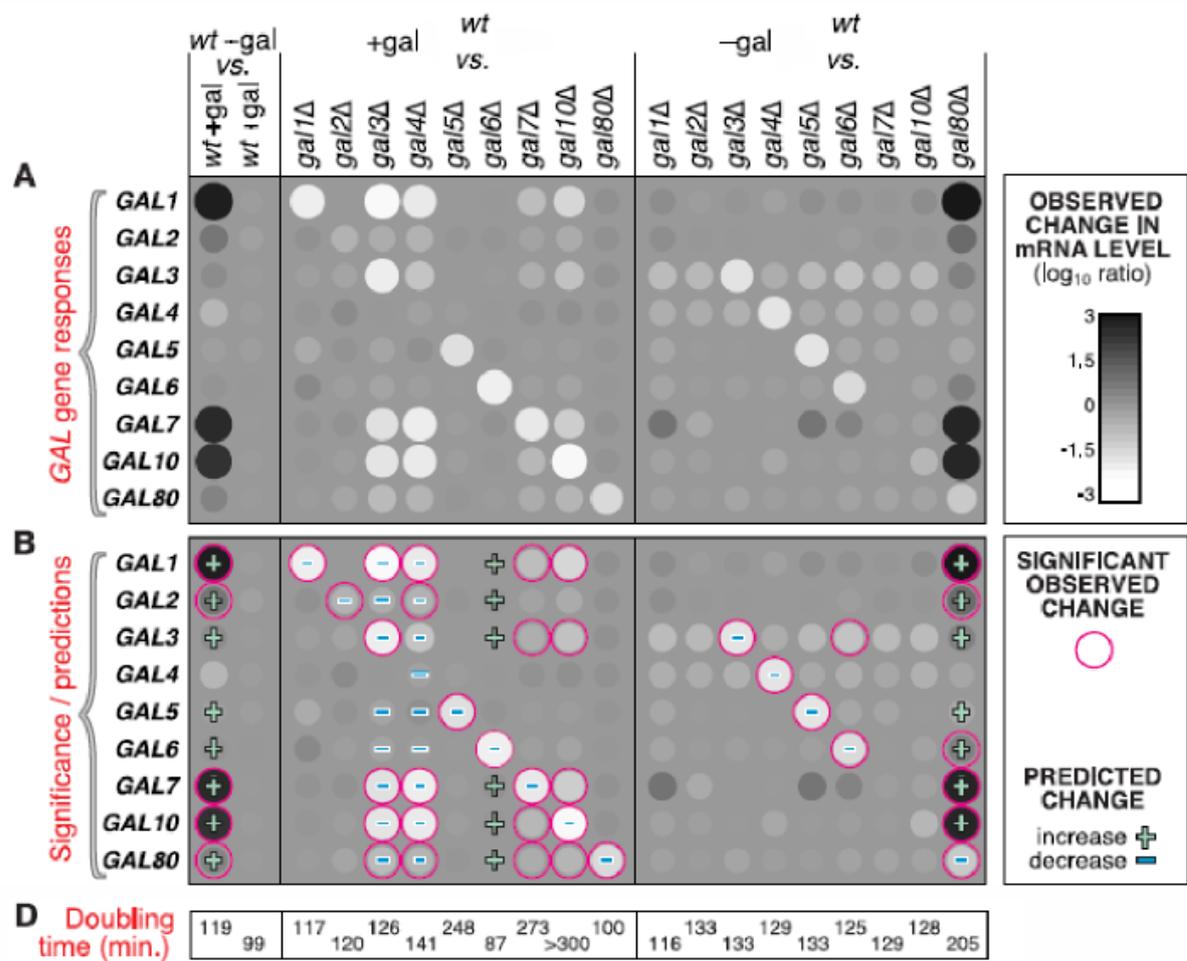


Abbildung 2: mRNA-Expressionsrate zwischen verschiedenen Hefestämmen im -gal und +gal Medium. (Abbildung leicht modifiziert nach [1])

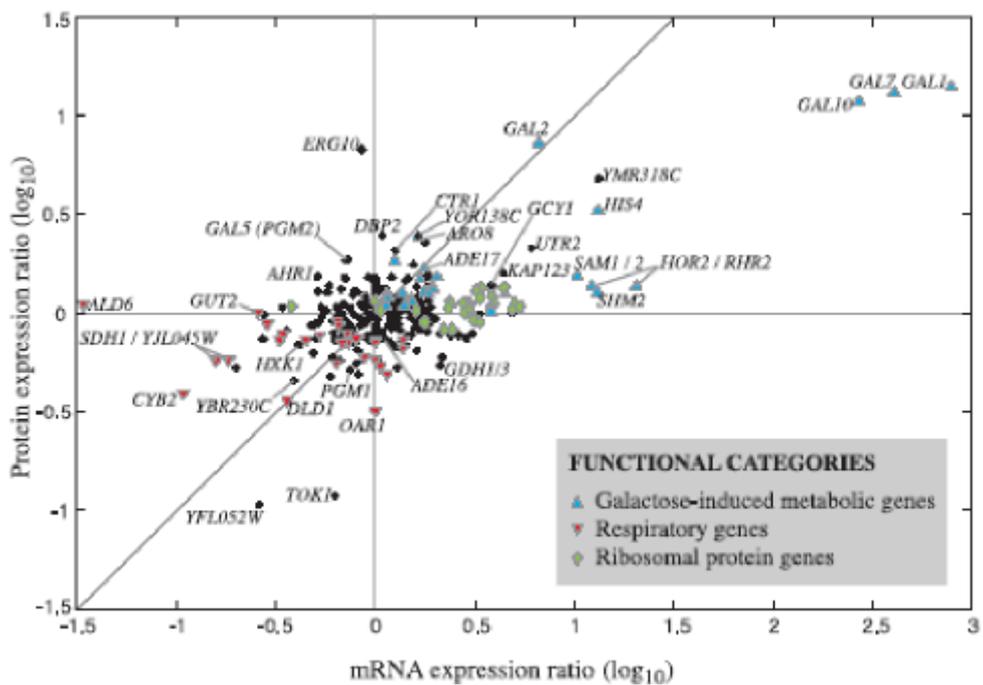


Abbildung 3: Protein vs. mRNA Expressionsraten zwischen dem -gal und dem +gal Wildtyp. Die Gene im GAL-System zeigen eine höhere Expression im +gal-Medium, die jedoch nicht gleich stark bei allen Genen ist. Auffallend sind auch die ribosomalen Proteine, die trotz erhöhter mRNA-Expression im +gal Medium eine relativ stabile Proteinexpression zeigen (Abbildung nach [1]).

In den untersuchten mutanten Stämmen wurden ausschließlich Gene des Galaktoseweges verändert. Wie aber die Experimente gezeigt haben, führten die Mutationen nicht nur im GAL-System zu Veränderungen, sondern betrafen die mRNA-Expression von fast 1000 Genen. Die Autoren entscheiden sich deshalb ein weitaus größeres Interaktionsnetzwerk zusammenzustellen, das es auch erlaubt, Interaktionen zwischen Komponenten des Galaktoseweges und anderen Subnetzen zu finden. Wie Abbildung 4 zeigt, handelt es sich um ein hochverzweigtes Netzwerk. Dieses physische Interaktionsnetzwerk wurde mit Hilfe von öffentlichen Datenbanken (u.a. [7]) und Resultaten aus wissenschaftlichen Veröffentlichungen erstellt. Eine Beobachtung der Autoren ist, dass die Gene mit einer korrelierten Expression in den mRNA-Messungen, ebenfalls an bekannten Interaktionen (z.B. Protein-DNA-Interaktion oder Protein-Protein-Interaktion) beteiligt sind. Auf diese Weise lassen sich Hypothesen darüber aufstellen, welche Gene einen Einfluss auf ihre Interaktionspartner unter verschiedenen Bedingungen haben. Um nun aber die Verzahnung des Galaktoseweges mit anderen Netzwerken genauer zu untersuchen, wurden Bindungsstellen für den zentralen Transkriptionsfaktor GAL4p in allen Upstreamregionen von Genen gesucht, die eine ähnliche mRNA-Expression aufwiesen. Für 16 dieser insgesamt 94 Gene wurde eine Bindungsstelle für GAL4p gefunden, wobei sieben davon bereits beschriebenen GAL-Gene im Galaktoseweg sind. Die restlichen neun Gene (MLF3, PCL10, YEL057C, YPL066W, YPS3, GAL3, RPA49, YMR318C, YPR194C) könnten nach Meinung der Autoren bisher unaufgedeckte Interaktionspartner und damit mögliche Ziele von GAL4p sein. Einige sind in Abbildung 1 eingezeichnet. Die Information der Proteinexpressionsmessung ist hier nur zum Teil eingeflossen. Leider lässt sich nur schwer einschätzen wie gut fundiert diese Hypothesen sind. Für eine genauere Betrachtung der Güte der Vorhersage aus verschiedenen Experimenten und Informationsquellen siehe Lu et al.[8] und Xia et al.[9].

Im letzten Schritt (iv) der Untersuchung widmen sich die Autoren der mRNA-Störungsmatrix und versuchen Beobachtungen zu erklären, die durch das initiale Modell nicht vorhergesagt wurden³. Eine Beobachtung ist die ungewöhnlich hohe Verdoppelungszeit der gal80Δ Mutante im -gal Medium und die ungewöhnlich hohe Zahl an Genen, die im Vergleich zum Wildtyp eine differentielle Expression zeigen. Die gal80Δ Mutante zeichnet sich dadurch aus, dass durch die fehlende Inhibition des Transkriptionsfaktors GAL4p alle von ihm regulierten GAL-Gene aktiviert werden, obwohl keine Galaktose im Medium vorhanden ist. Es liegt also die Vermutung nahe, dass die erhöhte Aktivität der GAL-Gene mit den Beobachtungen im Zusammenhang steht (z.B. durch Aktivierung weiterer Gene durch GAL-Proteine oder schlechte Energieversorgung der Zelle durch Anschalten unbenutzter Metabolismuswege, im speziellen dem Galaktoseweg). Für weitere Untersuchungen wurde eine gal4Δgal80Δ Doppelmutante auf ihre mRNA Expression untersucht, bei der die GAL-Enzyme und der Transporter GAL2 nicht exprimiert werden. Es zeigt sich, dass diese Doppelmutante ein ähnliches Expressionsmuster der Gene zeigt wie die gal4Δ Mutante. Das bedeutet, dass GAL80 nicht allein für den Effekt verantwortlich ist, sondern andere GAL-Gene eine Vermittlerrolle spielen. Um die Wirkung anderer GAL-Gene zu testen, wurde eine weitere Doppelmutante gal2Δgal80Δ untersucht, bei der die Effekte der hohen Verdoppelungszeit und der differentiellen Expression vieler anderer Gene nicht erkennbar waren. Das deutet darauf hin, dass der Transporter mit dem langsamen Wachstum und der global veränderten Genexpression in Verbindung steht, obwohl die Wirkung anderer GAL-Gene offen bleibt.

³eine vollständige Aufstellung findet sich unter <http://www.sciencemag.org/cgi/content/full/292/5518/929/DC1>

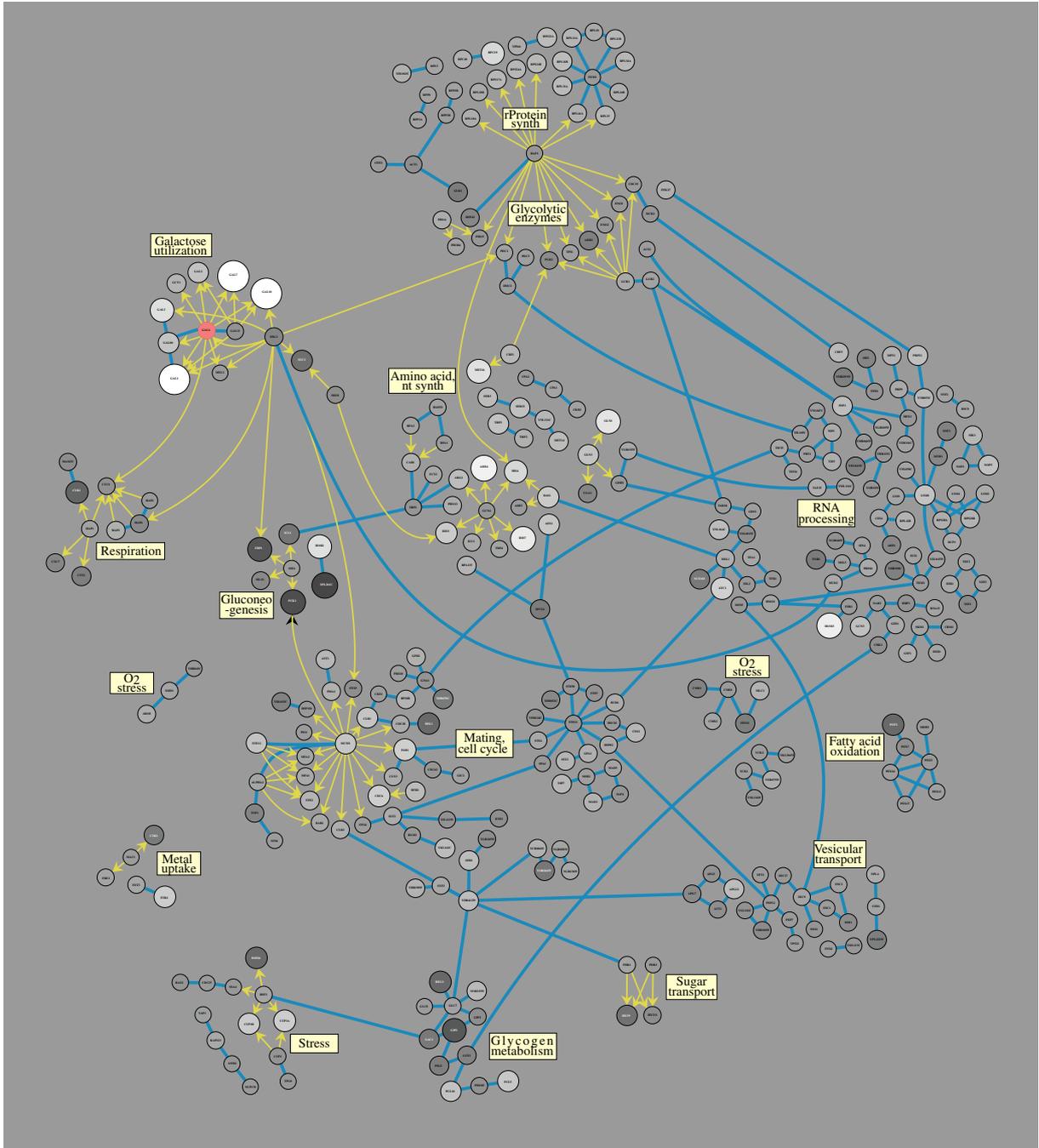


Abbildung 4: hochverzweigtes Interaktionsnetzwerk von Genen und ihren Produkten. Gelbe Pfeile stellen Protein-DNA Interaktionen dar, blaue Pfeile stehen für Protein-Protein Interaktionen. Die Gene/Proteine sind nach Funktion gruppiert (Abbildung nach [1]).

Die Arbeit der Autoren liefert wichtige Erkenntnisse über die Verwendbarkeit von mRNA und Proteindaten, auch wenn die vorgeschlagenen Erweiterungen des Modells bisher anscheinend nicht verifiziert werden konnten. Die Autoren erwähnen auch ausdrücklich, dass es sich hierbei nur um Hypothesen handelt, die noch weiterer Experimente bedürfen. Es ist ihnen jedoch gelungen zu demonstrieren, dass eine integrierte Analyse ein mächtiges Werkzeug für die Aufklärung von Netzwerken und Interaktionen ist, mit der zum Beispiel Hinweise auf eine posttranskriptionalen Kontrollmechanismus gefunden werden können. Am Beispiel wurde auch deutlich wie stark die Expression einzelner Gene die Expressionmuster anderer Gene beeinflussen kann. Die Bedeutung des Papers lässt sich nicht zuletzt an der Zahl der Zitierungen ablesen, die bis zum heutigen Datum nennenswerte 128 beträgt⁴. Nichtsdestotrotz bleiben einzelne Analyseverfahren, wie Microarrays, für sich allein ein wichtiges Werkzeug, z.B. für die Bestimmung von Zelltypen oder DNA-Markern.

Literatur

- [1] T Ideker, V Thorsson, J A Ranish, R Christmas, J Buhler, J K Eng, R Bumgarner, D R Goodlett, R Aebersold, and L Hood. Integrated genomic and proteomic analyses of a systematically perturbed metabolic network. *Science*, 292(5518):929–34, 2001.
- [2] S P Gygi, Y Rochon, B R Franza, and R Aebersold. Correlation between protein and mRNA abundance in yeast. *Mol Cell Biol*, 19(3):1720–30, 1999.
- [3] SG Gomes L Olsson J Nielsen. S Ostergaard, KO Walloe. The impact of gal6, gal80, and mig1 on glucose control of the gal system in saccharomyces cerevisiae.. *FEMS Yeast Res.*, 1(1):47–55, 2001.
- [4] A.P. Gasch et al. Genomic expression programs in the response of yeast cells to environmental changes. *Mol. Biol. Cell*, 11:4241–57, 2000.
- [5] Erez Braun and Naama Brenner. Transient responses and adaptation to steady state in a eukaryotic gene regulation system. *Phys Biol*, 1:67–76, 2004.
- [6] Timothy J Griffin, Steven P Gygi, Trey Ideker, Beate Rist, Jimmy Eng, Leroy Hood, and Ruedi Aebersold. Complementary profiling of gene expression at the transcriptome and proteome levels in *Saccharomyces cerevisiae*. *Mol Cell Proteomics*, 1(4):323–33, 2002.
- [7] V Matys, E Fricke, R Geffers, E Gossling, M Haubrock, R Hehl, K Hornischer, D Karas, A E Kel, O V Kel-Margoulis, D-U Kloos, S Land, B Lewicki-Potapov, H Michael, R Munch, I Reuter, S Rotert, H Saxel, M Scheer, S Thiele, and E Wingender. TRANSFAC: transcriptional regulation, from patterns to profiles. *Nucleic Acids Res*, 31(1):374–8, 2003.
- [8] Long J Lu, Yu Xia, Alberto Paccanaro, Haiyuan Yu, and Mark Gerstein. Assessing the limits of genomic data integration for predicting protein networks. *Genome Res*, 15(7):945–53, 2005.

⁴nach Science Magazine <http://www.sciencemag.org/cgi/content/abstract/292/5518/929>; Stand: 07.03.2006

- [9] Yu Xia, Haiyuan Yu, Ronald Jansen, Michael Seringhaus, Sarah Baxter, Dov Greenbaum, Hongyu Zhao, and Mark Gerstein. Analyzing cellular biochemistry in terms of molecular networks. *Annu Rev Biochem*, 73(NIL):1051–87, 2004.