

(12) NACH DEM VERTRAG ÜBER DIE INTERNATIONALE ZUSAMMENARBEIT AUF DEM GEBIET DES PATENTWESENS (PCT) VERÖFFENTLICHTE INTERNATIONALE ANMELDUNG

(19) Weltorganisation für geistiges Eigentum  
Internationales Büro



(43) Internationales Veröffentlichungsdatum  
17. Januar 2008 (17.01.2008)

PCT

(10) Internationale Veröffentlichungsnummer  
**WO 2008/006469 A1**

- (51) Internationale Patentklassifikation:  
**G06F 19/00** (2006.01)
- (21) Internationales Aktenzeichen: PCT/EP2007/005712
- (22) Internationales Anmeldedatum:  
28. Juni 2007 (28.06.2007)
- (25) Einreichungssprache: Deutsch
- (26) Veröffentlichungssprache: Deutsch
- (30) Angaben zur Priorität:  
10 2006 031 979.6 11. Juli 2006 (11.07.2006) DE
- (71) Anmelder (für alle Bestimmungsstaaten mit Ausnahme von US): **BAYER TECHNOLOGY SERVICES GMBH** [DE/DE]; 51368 Leverkusen (DE).
- (72) Erfinder; und
- (75) Erfinder/Anmelder (nur für US): **SCHUPPERT, Andreas** [DE/DE]; Stockberggasse 46, 51515 Kürten (DE). **ELLINGER-ZIEGELBAUER, Heidrun** [DE/DE]; Düppelstrasse 2a, 42781 Haan (DE). **AHR, Hans-Jürgen** [DE/DE]; Am Rohm 82, 42113 Wuppertal (DE).
- (74) Gemeinsamer Vertreter: **BAYER TECHNOLOGY SERVICES GMBH**; Law and Patents, Patents and Licensing, 51368 Leverkusen (DE).

- (81) Bestimmungsstaaten (soweit nicht anders angegeben, für jede verfügbare nationale Schutzrechtsart): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BH, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RS, RU, SC, SD, SE, SG, SK, SL, SM, SV, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.
- (84) Bestimmungsstaaten (soweit nicht anders angegeben, für jede verfügbare regionale Schutzrechtsart): ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), eurasisches (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), europäisches (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IS, IT, LT, LU, LV, MC, MT, NL, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Veröffentlicht:

— mit internationalem Recherchenbericht

[Fortsetzung auf der nächsten Seite]

(54) Title: METHOD FOR DETERMINING THE BEHAVIOR OF A BIOLOGICAL SYSTEM AFTER A REVERSIBLE DISTURBANCE

(54) Bezeichnung: VERFAHREN ZUR BESTIMMUNG DES VERHALTENS EINES BIOLOGISCHEN SYSTEMS NACH EINER REVERSIBLEN STÖRUNG

(57) Abstract: The invention relates to a method for determining the behavior of at least one biological system after a reversible disturbance comprising the following steps: (a) provision of at least one biological system, the biological system comprising a biological network consisting of a multiplicity of biological or biochemical components having an activity; (b) provision of a linear model for describing the behavior of the network of the biological system; (c) determination of the activity of the biological or biochemical components of the biological network; (d) reversible disturbance of the activity of at least one of the biological or biochemical components; (e) determination of the activity of the biological or biochemical components of the biological network after executing the reversible disturbance,- (f) determination of the change in the activity of at least one biological or biochemical component of the biological network as a response to the reversible disturbance; (g) calculation of the behavior of the biological network with the aid of the linear model provided for describing the behavior of the biological network and the change in activity determined in step (f) of the biological or biochemical component (n) after the reversible disturbance, taking into account the biodiversity of the response of the biological or biochemical component (n).

(57) Zusammenfassung: Die Erfindung betrifft ein Verfahren zur Bestimmung des Verhaltens wenigstens eines biologischen Systems nach einer reversiblen Störung umfassend die folgenden Schritte: (a) Bereitstellen wenigstens eines biologischen Systems, wobei das biologische System ein biologisches Netzwerk umfassend eine Vielzahl von biologischen oder biochemischen Komponenten, die eine Aktivität aufweisen, umfasst; (b) Bereitstellen eines linearen Modells zur Beschreibung des Verhaltens des Netzwerks des biologischen Systems; (c) Bestimmen der Aktivität der biologischen oder biochemischen Komponenten des biologischen Netzwerks; (d) reversibles Stören der Aktivität wenigstens einer der biologischen oder biochemischen Komponenten; (e) Bestimmen der Aktivität der biologischen oder biochemischen Komponenten des biologischen Netzwerks nach Ausüben der reversiblen Störung,- (f) Bestimmen der Änderung der Aktivität wenigstens einer biologischen oder biochemischen Komponente des biologischen Netzwerks als Reaktion auf die reversible Störung; (g) Berechnen des Verhaltens des biologischen Netzwerks anhand des bereitgestellten linearen Modells zur Beschreibung des Verhaltens des biologischen Netzwerks und der in Schritt (f) bestimmten Änderung der Aktivität der biologischen oder biochemischen Komponente (n) des biologischen Netzwerks nach der reversiblen Störung, unter Berücksichtigung der Biodiversität der Reaktion der biologischen oder biochemischen Komponente (n).



A1

WO 2008/006469



---

*Zur Erklärung der Zweibuchstaben-Codes und der anderen Abkürzungen wird auf die Erklärungen ("Guidance Notes on Codes and Abbreviations") am Anfang jeder regulären Ausgabe der PCT-Gazette verwiesen.*

### **Verfahren zur Bestimmung des Verhaltens eines biologischen Systems nach einer reversiblen Störung**

Die Erfindung betrifft ein Verfahren zur Bestimmung des Verhaltens wenigstens eines biologischen Systems nach einer reversiblen Störung.

5 Eukariotische wie auch prokariotische Zellen, die einem externen Stress ausgesetzt sind, zeigen signifikante Änderungen der Expression mehr oder weniger großer Gruppen von Genen, wobei bis zu 30 % aller Gene beeinflusst sein können. Hieraus lässt sich ableiten, dass eine Änderung der Genexpression als Antwort auf einen Stress weder ein lokales Phänomen in einem Netzwerk einander regulierender Gene darstellt, noch dass die Stressantwort auf isolierte Gene, Moleküle  
10 oder Signalwege beschränkt ist, selbst wenn die kausale Wirkungsweise des Stresses direkt lediglich wenige Gene betreffen mag. Zwischen verschiedenen Signalwegen besteht offensichtlich eine gegenseitige Beeinflussung und ein hoher Informationsaustausch, der es einer Zelle erlaubt, die zelluläre Stressantwort von ihrer lokalen Wirkung auf große Teile der Genexpression auszuweiten.

15 Die allgemeine Wirkung auf einen toxischen Stress auf Proteinebene wurde beispielsweise für Protein-Proteinwechselwirkung Netzwerk in *S. cerevisiae* und *E. coli*-Bakterien untersucht, wobei gezeigt werden konnte, dass ein toxischer Stress eine Stressantwort großer Gruppen von Proteinen hervorruft.

Es wird angenommen, dass die Organisationsstruktur der Stressantwort in Form sehr komplex  
20 wechselwirkender Hierarchien zu beschreiben ist, die ihrerseits auf lokalen Wechselwirkungen im Gesamtnetzwerk beruhen, die als biologische Signalwege und umfassende funktionelle Module interpretiert werden können. Die biologische Regulation eines Stresses kann somit umfassende Auswirkungen auf die Aktivität zellulärer Netzwerke haben und einen Austausch zwischen verschiedenen Signalwegen und funktionellen Einheiten beinhalten.

25 Eine globale Modulation der Genexpression legt nahe, dass ein ganzheitlicher Ansatz auf der Basis von generischen Eigenschaften von großräumigen Mechanismen der Stressantwort in Netzwerken geeignet sein könnte, eine solche Stressantwort zu beschreiben.

Im Stand der Technik sind Verfahren zur Bestimmung einer solchen Stressantwort bekannt. Beispielsweise offenbart die Schrift WO 03/077062 sowie in "Gardner et al., Science, Vol. 301  
30 (5629), S. 102-5 (4. Juli 2003)" ein Modell zur Beschreibung einer stressinduzierten Änderung der Genexpression unter Verwendung einer Gruppe von Differenzialgleichungen, die die Aktivität der einzelnen Elemente des Netzwerks durch Variablen wiedergeben. Nachteilig bei diesem Modell ist

jedoch, dass die die Gleichungen quantifizierende Matrix, die die Interaktionen der einzelnen Elemente beschreibt, explizit berechnet werden muss. Voraussetzung für die explizite Berechnung der Interaktion der einzelnen Elemente ist, dass die Interaktion der einzelnen Elemente bekannt ist. Dies ist beispielsweise für Gene in den wenigstens Fällen ausreichend bekannt. Dann beinhaltet  
5 eine solche Berechnung, dass experimentell über exakt definierte Störungen die Wechselwirkung der einzelnen Komponenten ermittelt werden muss. Für eine größere Anzahl an Elementen ist eine explizite Berechnung mit diesem Model somit nicht möglich und das beschreibbare Netzwerk ist auf eine sehr geringe Zahl an Elementen und deren Wechselwirkungen beschränkt.

Der Erfindung lag entsprechend die Aufgabe zugrunde, ein Modell zur Beschreibung von  
10 Änderungen der Genexpression als Antwort auf einen externen Stress zur Verfügung zu stellen, das die vorgenannten Nachteile des Standes der Technik überwindet. Insbesondere war die Aufgabe der vorliegenden Erfindung ein Verfahren zur Verfügung zu stellen, das eine Bestimmung einer Stressantwort in Netzwerken ermöglicht, ohne dass die explizite Wechselwirkung der Elemente bekannt sein muss.

15 Erfindungsgemäß wird die Aufgabe dadurch gelöst, dass ein Verfahren zur Bestimmung des Verhaltens wenigstens eines biologischen Systems nach einer reversiblen Störung zur Verfügung gestellt wird, das die folgenden Schritte umfasst:

- (a) Bereitstellen wenigstens eines biologischen Systems, wobei das biologische System ein biologisches Netzwerk umfassend eine Vielzahl von biologischen oder biochemischen  
20 Komponenten, die eine Aktivität aufweisen, umfasst;
- (b) Bereitstellen eines linearen Modells zur Beschreibung des Verhaltens des Netzwerks des biologischen Systems;
- (c) Bestimmen der Aktivität der biologischen oder biochemischen Komponenten des biologisches Netzwerks;
- 25 (d) reversibles Stören der Aktivität wenigstens einer der biologischen oder biochemischen Komponenten, wobei eine Reaktion des biologisches Netzwerks erzeugt wird, die durch die Änderung der Aktivität wenigstens einer oder mehrerer der biologischen oder biochemischen Komponenten ausgebildet wird;
- (e) Bestimmen der Aktivität der biologischen oder biochemischen Komponenten des  
30 biologisches Netzwerks nach Ausüben der reversiblen Störung, sobald die Komponenten des Netzwerks die Reaktion auf die Störung vollzogen haben;

- (f) Bestimmen der Änderung der Aktivität wenigstens einer biologischen oder biochemischen Komponente des biologischen Netzwerks als Reaktion auf die reversible Störung;
- (g) Berechnen des Verhaltens des biologischen Netzwerks anhand des bereitgestellten linearen Modells zur Beschreibung des Verhaltens des biologischen Netzwerks und der in Schritt (f) bestimmten Änderung der Aktivität der biologischen oder biochemischen Komponente(n) des biologischen Netzwerks nach der reversiblen Störung, unter Berücksichtigung der Biodiversität der Reaktion der biologischen oder biochemischen Komponente(n); und
- (h) optional Vergleichen zwischen der gemäß Schritt (f) bestimmten Änderung der Aktivität der einzelnen Komponenten und dem gemäß Schritt (g) berechneten Verhalten des biologischen Netzwerks anhand des bereitgestellten linearen Modells, wobei man erwartet, dass eine Übereinstimmung des berechneten Verhaltens mit den in Schritt (f) bestimmten Änderung der Aktivität der biologischen oder biochemischen Komponente(n) besteht.

Weitere Gegenstände der vorliegenden Erfindung betreffen ein Computerprogrammprodukt, ein Computerprogramm und ein Computersystem zur Ausführung eines oder mehrerer Schritte des erfindungsgemäßen Verfahrens.

Weitere vorteilhafte Ausgestaltungen der Erfindung ergeben sich aus den Unteransprüchen.

Unter dem Begriff "biologisches System" wird im Sinne der vorliegenden Erfindung eine Zelle oder eine Zellpopulation, beispielsweise ein Gewebe oder ein Organ, wie die Leber, oder ein multizellulärer Organismen, insbesondere ein Säugetier wie Maus oder Ratte, verstanden. In bevorzugten Ausführungsformen ist das biologische System ausgewählt aus der Gruppe umfassend Zelle(n), Gewebe, Organ(e) und/oder Organismen.

Ein biologisches System enthält eine Vielzahl von biologischen oder biochemischen Komponenten. Unter dem Begriff "biologische Komponente" werden im Sinne der vorliegenden Erfindung biologische zelluläre Bestandteile verschiedener Art verstanden, beispielsweise Gene, die miteinander in Verbindung stehen und/oder einander beeinflussen können. Es versteht sich, dass die Art der biologischen Komponente abhängt von der Art des betrachteten biologischen Systems. Ist das betrachtete biologische System eine Zelle, so sind die biologischen Komponenten ausgewählt aus der Gruppe zelluläre Bestandteile, insbesondere Gene. Ist das betrachtete biologische System eine Zellpopulation wie Gewebe oder Organ, so können die biologischen Komponenten Gene wie auch einzelne Zellen sein.

Unter dem Begriff "biochemische Komponente" werden im Sinne der vorliegenden Erfindung biochemische zelluläre Bestandteile verschiedener Art verstanden, beispielsweise Moleküle, die mit-

einander in Verbindung stehen und/oder einander beeinflussen können. In bevorzugten Ausführungsformen ist die biochemische Komponente ausgewählt aus der Gruppe umfassend in der Zelle oder Zellpopulationen enthaltenen Molekülen wie Desoxyribonukleinsäure (DNA, DNS), Ribonukleinsäure (RNA, RNS), Proteine und/oder Metabolite.

- 5 Unter dem Begriff "Aktivität" wird im Sinne der vorliegenden Erfindung verstanden, dass eine biologische oder biochemische Komponente eine Eigenschaft oder Funktion aufweist. Beispielsweise werden Gene oder Proteine entweder exprimiert oder nicht exprimiert oder weisen eine Expressionsrate auf, die beispielsweise als Gehalt an RNA oder Genprodukt bestimmbar ist. Weiterhin können Gene oder Proteine in bestimmter Menge oder Konzentration vorliegen, Funktionen ausüben, beispielsweise katalytische Wirkungen, die durch chemische Modifizierung des  
10 Gens oder Proteins variiert werden kann. Einer Aktivität oder dem Zustand einer Aktivität kann die Menge, Konzentration, Expressionsrate oder katalytische Funktion entsprechen. Eine chemische Modifizierung oder Funktionalisierung einer Komponente, beispielsweise eines Gens oder Proteins, kann einem Aktivitätszustand entsprechend, es kann im Rahmen der Erfindung  
15 jedoch auch vorgesehen sein, dass eine chemische Modifizierung oder Funktionalisierung zwei verschiedene biologische oder biochemische Komponenten definieren kann.

Unter dem Begriff "biologisches Netzwerk" wird im Sinne der vorliegenden Erfindung eine Gruppe oder Vielzahl von biologischen oder biochemischen Komponenten verstanden, die einander beeinflussen und/oder Auswirkungen auf die Aktivität anderer Komponenten haben  
20 können. Ein biologisches Netzwerk enthält vorzugsweise biologische oder biochemische Komponenten einer Art, es kann jedoch auch vorgesehen sein, dass ein biologisches Netzwerk biologische und/oder biochemische Komponenten verschiedener Art enthalten kann, die einander beeinflussen können. Beispielsweise kann ein biologisches Netzwerk Gene, RNA-Moleküle, Proteine und/oder Metaboliten aufweisen, die einander gegenseitig in ihrer jeweiligen Aktivität beeinflussen können.

- 25 Unter dem Begriff "reversible Störung" wird im Sinne der vorliegenden Erfindung verstanden, dass die biologischen oder biochemischen Komponenten, das biologische Netzwerk und/oder das biologische System beeinflusst werden, wobei eine Störung insbesondere ein Stress sein kann, der auf das System wirkt. Insbesondere kann der Stress ein externer Stress sein, der von außen auf das System wirkt. Vorzugsweise ist ein Stress ausgewählt aus der Gruppe umfassend toxischer Stresse,  
30 bevorzugt Stress durch nicht-genotoxische oder genotoxische Hepatokarzinogene, Stress durch Applikation eines pharmazeutischen Wirkstoffs, Wärmestress oder Hunger. Ein Stress, der eine Störung des Systems verursacht, kann ebenfalls ein Wirkstoff und/oder ein Medikament sein, das dem System beigebracht wird. Eine Störung oder Stress ist dann reversibel, wenn das System nach Entfernen der Störung oder des Stresses in seinen Ausgangszustand zurückkehrt.

Eine Störung ruft im Sinne der Erfindung eine "Reaktion" der biologischen oder biochemischen Komponenten hervor. Unter dem Begriff "Reaktion" wird im Sinne der vorliegenden Erfindung verstanden, dass die Aktivität wenigstens einer der biologischen oder biochemischen Komponenten durch die Störung verändert wird. Beispielsweise kann durch eine Störung die Aktivität  
5 wenigstens einer biologischen oder biochemischen Komponenten geändert werden. Diese Änderung der wenigstens einen biologischen oder biochemischen Komponente kann wiederum die Aktivität wenigstens einer weiteren biologischen oder biochemischen Komponente beeinflussen. Eine Störung kann durch direkte oder indirekte Beeinflussung der biologischen oder biochemischen Komponenten eines biologischen Systems eine Reaktion einer, mehrerer oder einer  
10 Vielzahl der biologischen oder biochemischen Komponenten hervorrufen. Diese Reaktion der Komponenten bildet die Reaktion des Netzwerks, die entsprechend durch die Reaktion wenigstens einer, mehrerer oder vieler der biologischen oder biochemischen Komponenten ausgebildet wird.

Beispielsweise kann ein Wirkstoff nur die Aktivität eines Proteins beeinflussen oder die Konzentration eines Metaboliten erhöhen. Ein toxischer Stress kann beispielsweise viele verschiedene  
15 Gene direkt und indirekt in ihrer Aktivität beeinflussen und eine großräumige Stressantwort hervorrufen.

Unter dem Begriff "Verhalten des biologischen Netzwerks" wird im Sinne der vorliegenden Erfindung verstanden, dass das biologische Netzwerk auf die Änderung der Aktivität wenigstens einer der biologischen oder biochemischen Komponenten reagiert, indem die gegenseitige Beeinflussung der Komponenten Auswirkungen auf die Aktivität anderer Komponenten hat und das  
20 Netzwerk durch die Reaktionen der einzelnen Komponenten insgesamt seine Aktivität ändert. Beispielsweise kann ein Gen als Reaktion auf einen Stress seine Expression ändern, wobei die Expressionsänderung dieses Gens die Expression eines oder mehrerer anderer Gene beeinflusst, die untereinander oder bei weiteren Gene ebenfalls Expressionsänderungen bewirken können. Als  
25 Folge davon wird das Netzwerk miteinander korrespondierender Gene insgesamt eine Änderung bzw. Verschiebung der Expression erfahren.

Unter dem Begriff "Rauschen" wird im Sinne der vorliegenden Erfindung verstanden, dass die Reaktion der biologischen oder biochemischen Komponenten auf eine identische externe Störung oder Stress nicht identisch sein muss, sondern insbesondere bei biologischen Systemen eine  
30 Variation aufweist. Diese Variation kann beispielsweise graduelle Unterschiede in der Änderung der Expression eines Gens oder Proteins auf einen identischen Stressfaktor unter identischen Bedingungen hervorrufen. Diese Variation oder "Rauschen" der Reaktion der biologischen oder biochemischen Komponenten weist einen Rauschanteil, der auf Messrauschen und Messfehlern beruht, wie sie bei Experimenten regelmäßig auftreten, und einen biologischen Anteil auf, der im

Sinne der vorliegenden Erfindung als "Biodiversität" bezeichnet wird. Ein Rauschen kann insbesondere eine Fluktuationen der Gen- oder Proteinexpression sein. Das "Rauschen" der Gen- und Proteinexpression aufgrund von Biodiversität ist beispielsweise beschrieben in Bar-Even et al., Nature Genetics, Vol. 38, Nr. 6, S. 636-643, 2006, auf die Bezug genommen wird.

5 Unter dem Begriff "Biodiversität" wird im Sinne der vorliegenden Erfindung eine biologische Variation verstanden. Biodiversität kann biologische Variationen ausgewählt aus der Gruppe umfassend natürliche Variation einer Aktivität einer Komponente oder eines Netzwerks, eine natürliche Variation eines biologischen Systems und/oder eine Variation der biologischen Reaktionen eines Systems auf Umgebungsfaktoren sein. Beispielsweise kann der Begriff  
10 "Biodiversität" in dem biologischen System einer Zelle oder eines Gewebes umfassend ein Netzwerk aus vielen einzelnen Genen eine natürliche Variation der Genexpression eines einzelnen Gens, mehrerer Gene und/oder eines Netzwerks von Genen, oder eine natürliche Variation der Proteinexpression eines einzelnen Proteins, mehrerer Proteine und/oder eines Netzwerks von Proteinen in einem Proteinnetzwerk umfassen. Weiterhin kann eine "Biodiversität" bei einem  
15 Vergleich verschiedener biologischer Systeme, beispielsweise verschiedener Organismen einer Spezies, Variationen ausgewählt aus der Gruppe umfassend eine Variation des Genotyps, eine Variation einzelner Organe und/oder eine unterschiedliche Reaktion des Organismus auf äußere Einflüsse wie Ernährung umfassen. Es versteht sich, dass die Biodiversität die Aktivität oder Reaktionen der Komponenten, Netzwerke und/oder Systeme einander beeinflusst, so dass die  
20 Biodiversität der Reaktion der Komponenten auf eine Störung sowohl auf der natürlichen Variation der Aktivität der Komponenten beruhen kann wie auch eine natürlichen Variation eines biologischen Systems und/oder einer Variation der biologischen Reaktionen eines Systems auf Umgebungsfaktoren umfassen kann.

Der Begriff "Biomarker" wird als indirektes Beobachtungsverfahren für eine große Zahl intra- and  
25 extrazellulärer Ereignisse wie auch physiologische Änderungen eines Organismus verwendet, die nicht oder nur mit großem Aufwand direkt beobachtet werden können.. Hierzu kann beispielsweise des Gehalts oder der Produktionsrate von Signalmolekülen, Transkriptionsfaktoren, Metaboliten, Gentranskripten oder Modifikationen von Proteinen nach der Translation, oder auch der physiologische Zustand eines biologischen Systems gehören. Unter dem Begriff "Biomarker" wird im  
30 Sinne der vorliegenden Erfindung insbesondere eine Kombination aus einem Gen, oder Genprodukt, ein Protein, oder eine Gruppe von Genen, Genprodukten oder Proteinen verstanden, das nach einer Störung verglichen mit der Aktivität vor der Störung auf- oder abreguliert wird, und ein entsprechendes Berechnungsverfahren zur Berechnung nicht direkt beobachtbarer Größen. Für einen Biomarker ist insbesondere eine biologische oder biochemische Komponente oder eine  
35 Gruppe davon, ein Gen oder eine Gruppe von Genen, das ausreichend spezifisch auf eine spezielle

Störung reagiert wesentlich, so dass es allein oder in Kombination mit anderen Genen oder Genprodukten verwendbar ist, eine Klassifizierung von Störungen in Klassen, beispielsweise in Toxizitätsklassen, zu erlauben. Insbesondere ist ein Biomarker eine Kombination aus einer biologische oder biochemische Komponente oder eine Gruppe davon, ein Gen, eine Gruppe von Genen oder ein Genprodukt, die charakteristisch für eine Reaktion eines biologischen Systems auf eine bestimmte Störung ist, und einem zugehörigen Berechnungsverfahren.

Die Störung des Grundzustands der Aktivität bildet die Basis von Krankheiten, die mit einer Reaktion der Komponenten bzw. des Systems auf die Störung verbunden sind. Die vorliegende Erfindung beruht insbesondere auf der Hypothese, dass Störungen an beispielsweise toxischen Phänomenen beteiligt sein können und dass Biomarker d.h. eine oder mehrere Komponenten, die eine für die Reaktion des Systems charakteristische Aktivitätsänderung zeigen, wirksame Marker der Toxizität bilden könnten.

Ein Vorteil der vorliegenden Erfindung ist, dass dadurch, dass die Berechnung innerhalb des bereitgestellten linearen Modells zur Beschreibung des Verhaltens eines biologischen Netzwerks unter Berücksichtigung der Biodiversität der Reaktion der biologischen oder biochemischen Komponenten erfolgt, das Verhalten des biologischen Netzwerks berechenbar ist, ohne dass die Wechselwirkung aller Komponenten explizit berechnet werden muss. Von besonderem Vorteil ist hierbei, dass das Verhalten des Netzwerks aus bestimmbar oder messbaren Daten der Einzelreaktionen der Komponenten rekonstruierbar ist. Vorteilhafter Weise kann die Berechnung des Verhaltens des Netzwerks auf die Reaktionen der Komponenten auf die Störung und somit beobachtbare Größen zurückgeführt werden.

Insbesondere ist von großem Vorteil, dass es aufgrund der Berücksichtigung der Biodiversität erzeugten Variation der Reaktion der biologischen oder biochemischen Komponenten in dem Verfügung gestellten linearen Modell ermöglicht wird, das Verhalten des Netzwerks ohne systematische Experimente zu bestimmen.

Biologische Netzwerke können mathematisch dargestellt werden. Das im Rahmen des erfindungsgemäßen Verfahrens zur Verfügung gestellte lineare Modell zur Beschreibung des Verhaltens des biologischen Netzwerks umfasst eine mathematische Beschreibung der Reaktion biologischer oder biochemischer Komponenten eines Netzwerks auf eine reversible Störung. Eine reversible Störung, bei der das System wieder in seinen Ausgangszustand zurückkehrt, wenn der Stress entfernt wird, stört die Aktivität der Komponenten und führt zu einer Aktivitätsänderung der durch die Störung betroffenen Komponenten. Eine solche Änderung der Aktivität einer Komponente kann ihrerseits Auswirkung auf die Aktivität weiterer biologischer oder biochemischer Komponenten zeigen. Biologische und/oder biochemische Komponenten, die Komponenten eines biolo-

gischen Netzwerks sind, können miteinander wechselwirken und einander in ihrer Aktivität regulieren. Die Regulation kann positiv oder negativ sein, beispielsweise eine Hoch- oder Runterregulation der Genexpression für den Fall, dass die Komponenten Gene sind oder eine Hoch- oder Runterregulation der Proteinexpression für den Fall, dass die Komponenten Proteine sind. Ein  
 5 reversibles Stören der Aktivität wenigstens einer biologischen oder biochemischen Komponente erzeugt somit eine Reaktion der Komponenten des biologischen Netzwerks, die insgesamt die Reaktion des gesamten Netzwerks der Komponenten ausbilden.

Die Wechselwirkung oder Interaktion der einzelnen Komponenten in einem Netzwerk miteinander ist nicht zwangsläufig homogen. Einwertige Parameter können die Wechselwirkung der  
 10 Komponenten daher nicht beschreiben und ein generischer Formalismus zur Berechnung des Verhaltens eines biologischen Netzwerks ist im Sinne dieser Erfindung vorzugsweise geeignet.

Eine bevorzugte generische Beschreibung kann das bereitgestellte lineare Modell zur Beschreibung des Verhaltens des biologischen Netzwerks gemäß der nachstehenden Gleichung (I) zur Verfügung stellen:

$$15 \quad x = A u \quad (I)$$

worin:

x:  $[x_1 \dots x_n]$  ein Vektor ist, der die Bestimmung der Änderung der Aktivität wenigstens einer biologischen oder biochemischen Komponente des biologischen Netzwerks als Reaktion auf die reversible Störung umfasst,

20 u:  $[u_1 \dots u_n]$  ist ein Vektor, der die Störung beschreibt,

A: ist eine Matrix, die die Parameter enthält, die die Reaktion der Komponenten auf die Störung beschreiben,

n ist die Anzahl der Komponenten.

Vorzugsweise beschreibt man die Matrix durch eine symmetrische  $n \times n$ -Matrix, wobei n der  
 25 Anzahl der Komponenten entspricht. Diese enthält die Bestandteile  $a_{ij}$ , die quantitativ die Reaktion einer Komponente i auf einen Stress  $u_j$  beschreiben, der auf die Komponente j wirkt. Die Matrix A gibt also sowohl die Reaktion der Komponenten des Netzwerks auf die reversible Störung wieder als auch die Verteilung der Reaktion auf eine lokale Störung oder einen lokalen Stress, der nur auf nur wenige Komponenten wirkt, auf das ganze Netzwerk.

Der Vektor  $x$ , der die Änderung der Aktivität der einzelnen Komponenten angibt, gibt geeigneterweise Daten oder Messwerte wieder, die die Änderung der Aktivität der Komponenten nach einer reversiblen Störung beschreiben, nachdem die Komponenten des Netzwerks auf die Störung reagiert haben.

- 5 Die Komponenten reagieren abhängig von der Art der Komponente, beispielsweise Gene und/oder Proteine, und von der ausgeübten reversiblen Störung innerhalb unterschiedlicher Zeiträume auf die Störung mit einer Änderung ihrer Aktivität, wobei die Zeiträume einer reversiblen Reaktion der Komponenten können hierbei im Bereich von Minuten, Stunden oder Tagen liegen können. Diese Zeiträume sind dem Fachmann bekannt und/oder bestimmbar. Vorzugsweise sind im Schritt (f)
- 10 schnelle Reaktionen der Komponenten bestimmbar, beispielsweise Änderungen der Genexpression, die vorzugsweise im Bereich von 0,5 Stunden bis 24 Stunden nach dem Ausüben einer reversiblen Störung auftreten.

Die Größe der Matrix  $A$  hängt ab von der Anzahl an biologischen oder biochemischen Komponenten des Netzwerks. Diese Zahl kann in biologischen Netzwerken und/oder Systeme in

15 weiten Bereichen variieren. Ist das biologische System beispielsweise eine Zelle und die Komponenten sind Gene, können mehrere tausend Gene in einem Netzwerk enthalten sein. Die Größe eines solchen Netzwerks kann ebenfalls abhängig sein von der Störung, die auf das System einwirkt. Ist eine solche Störung beispielsweise ein toxischer Stress, können mehrere tausend Gene durch einen solchen Stress betroffen sein.

- 20 Anzahl der Komponenten  $n$  kann im Bereich von  $\geq 1$  Komponente bis  $\leq 25.000$  Komponenten, vorzugsweise im Bereich von  $\geq 1$  Komponente bis  $\leq 15.000$  Komponenten, bevorzugt im Bereich von  $\geq 1$  Komponente bis  $\leq 5000$  Komponenten, besonders bevorzugt im Bereich von  $\geq 2$  Komponenten bis  $\leq 1000$  Komponenten, weiter vorzugsweise im Bereich von  $\geq 5$  Komponenten bis  $\leq 400$  Komponenten, noch vorzugsweise im Bereich von  $\geq 5$  Komponenten bis  $\leq 200$
- 25 Komponenten liegen.

In bevorzugten Ausführungsformen des erfindungsgemäßen Verfahrens werden die Eigenschaften der Matrix  $A$  aus der bestimmaren Änderung der Aktivität der biologischen oder biochemischen Komponenten des biologischen Netzwerks nach der reversiblen Störung unter Berücksichtigung der Biodiversität der Reaktion der Komponenten beschrieben. Eine solche Berechnung erfolgt

30 vorzugsweise dadurch, dass der Vektor  $u$ , der die Störung beschreibt, die auf die Komponenten wirkt, einen Rauschanteil aufweist, der das Messrauschen wiedergibt, das beispielsweise durch Messungenauigkeiten und/oder Messfehler verursacht werden kann, und einen Rauschanteil, der das Rauschen aufgrund der Biodiversität der Reaktion der Komponenten wiedergibt, wobei die Biodiversität dieser Reaktion die biologische Variation der Reaktion der Komponenten wiedergibt.

Das bereitgestellte lineare Modell zur Beschreibung des Verhaltens des Netzwerks umfassend die Matrix A ist eine lineare Näherung eines nicht linearen Systems. Eine solche lineare Näherung des Verhaltens des Netzwerks ist einem grundsätzlich nicht linearen System dann äquivalent, wenn sich das System in oder nahe einem stationären Zustand befindet. Ist das biologische System  
5 beispielsweise eine Zelle, eine Zellkultur oder ein Organismus, beispielsweise eine Ratte, bedeutet dies, dass Zellen oder Organismen vorzugsweise in einer konstanten Umgebung zu halten sind.

Im Rahmen dieses erfindungsgemäßen Verfahrens wird weiterhin eine reversible Störung bevorzugt ein reversibler Stress auf das System ausgeübt, wobei das System nach Entfernen der Störung oder des Stresses in den Ausgangszustand zurückkehrt. Eine solche reversible Störung  
10 erlaubt entsprechend die Anwendung eines linearen Modells zur Beschreibung des Verhaltens des Netzwerks. Die Rückkehr des Systems in den Ausgangszustand umfasst regelmäßig ein sogenanntes Rauschen, das im Sinne der vorliegenden Erfindung so verstanden wird, dass die Reaktion der biologischen oder biochemischen Komponenten auf eine identische Störung oder Stress nicht identisch sein muss, sondern eine Variation aufweist. Diese Variation bedingt, dass die Kompo-  
15 nenten den Ausgangszustand erreichen können oder sich dem Ausgangszustand nähern können, wobei der Zustand den das System bzw. die einzelne Komponente nach der Störung einnimmt, ihrem Ausgangszustand entspricht, der durch das Rauschen überlagert ist.

Dieses Rauschen oder Variationen der Reaktion der biologischen oder biochemischen Komponenten und/oder des biologischen Systems lässt sich in einen Rauschanteil unterteilen, der auf  
20 Messrauschen und/oder Messfehlern beruht und einen biologischen Rauschanteil, der auf der biologischen Variation der Komponenten und/oder des Systems beruht und im Sinne der vorliegenden Erfindung als Biodiversität bezeichnet wird.

Ist die Komponente beispielsweise ein Gen und das biologische System ein Gewebe oder eine Zelle, dem ein Stress zugefügt wird, führt das Rauschen dazu, dass die Expression eines Gens,  
25 nachdem sie sich als Reaktion auf die reversible Störung geändert hat, nach Ende der Störung ihren Ausgangswert nicht exakt wieder einnehmen muss, sondern um den Ausgangswert variieren kann. Auch bei einer oder mehreren Wiederholungen beispielsweise in wenigstens einem identischen System und/oder mit wenigstens einer identischen Störung oder Stress, wird die Komponente bzw. das System nach einer reversiblen Störung in den Ausgangszustand zurückkehren bzw.  
30 einen Zustand einnehmen, der eine Variation oder Streuung um den Ausgangszustand aufweist.

Es wird für die Anwendung des Modells zur Voraussage des Verhaltens des Netzwerks vorausgesetzt, dass sich das System sich in einem stationären Zustand oder "steady state" befindet. Das Ausüben einer reversiblen Störung oder eines reversiblen Stresses führt dazu, dass das System

nach Entfernen der Störung oder des Stresses in diesen stationären Ausgangszustand bis auf Abweichungen, erzeugt durch die Biodiversität, zurückkehrt.

Gemäß dem erfindungsgemäßen Verfahren bestimmt man in Schritt (c) die Aktivität der biologischen oder biochemischen Komponenten des biologisches Netzwerks im Ausgangszustand, stört  
5 gemäß Schritt (d) reversible die Aktivität wenigstens einer der biologischen oder biochemischen Komponenten, wobei eine Reaktion des biologisches Netzwerks erzeugt wird, die durch die Änderung der Aktivität wenigstens einer oder mehrerer der biologischen oder biochemischen Komponenten ausgebildet wird, und bestimmt gemäß Schritt (e) die Aktivität der biologischen oder biochemischen Komponenten des biologisches Netzwerks nach Ausüben der reversiblen Störung,  
10 sobald die Komponenten des Netzwerks die Reaktion auf die Störung vollzogen haben.

Vorteilhafter Weise ist eine Wiederholung des erfindungsgemäßen Verfahrens zur Voraussage des Verhaltens des Netzwerks nicht notwendig. Somit liegt ein besonderer Vorteil des Verfahrens darin, dass eine Berechnung ermöglicht wird, durch eine Messung nach einer Störung in einem System und der Ausgangszustand bekannt ist oder bestimmt wird.

15 Unter der Berücksichtigung der Biodiversität der Reaktion der biologischen oder biochemischen Komponenten weist der Vektor  $u$ , der die auf jede Komponente wirkende Störung beschreibt, einen Anteil auf, der das Messrauschen wiedergibt und einen Anteil, der die biologische Variation oder Biodiversität wiedergibt. Wird der Anteil des Messrauschens als konstanter Faktor betrachtet, kann die Reaktion auf eine Störung als auf die Biodiversität beschränkt angenommen werden.  
20 Weiterhin kann angenommen werden, dass die Biodiversität oder der biologische Anteil des Rauschens eine energetische Gleichverteilung aufweist und in Bezug auf die einzelnen Parameter  $u_1$  bis  $u_n$  eine Gleichverteilung aufweist. Die einzelnen Parameter  $u_1$  bis  $u_n$  werden auch als Anregungsmoden bezeichnet.

In bevorzugten Ausführungsformen des Verfahrens wird die Matrix durch eine Projektion der  
25 Daten der Änderung der bestimmten Aktivität der Komponenten auf ihre Eigenvektoren mit Hilfe der Korrelationskoeffizienten von Komponentenpaaren des biologischen Netzwerks beschrieben.

Die Eigenvektoren der Matrix  $A$  beschreiben formal Komponentengruppen des Netzwerks, die sich in ihrer Reaktion auf eine Störung oder Stress kohärent verhalten. Der zugehörige Eigenwert beschreibt die Sensitivität auf eine Störung oder Stress der jeweiligen Komponentengruppe bei  
30 kohärentem Reaktionsverhalten.

Die Korrelationskoeffizienten der Komponentenpaare des biologischen Netzwerks sind in Form der Eigenwerte und Eigenvektoren der Matrix  $A$  bestimmbar. Die Eigenwerte sind erhältlich aus

der Biodiversität der Reaktion der Komponenten unter der Annahme, dass die Biodiversität einem thermischen Rauschen entspricht. Unter dieser Voraussetzung kann das Reaktionsverhalten des Netzwerks respektive die relevanten Eigenvektoren der Matrix A aus eine Analyse des Rauschverhaltens berechnet werden.

- 5 Die Matrix A sei vorzugsweise eine elastische Matrix. Hierbei sei  $\{\lambda_i^*\}$  die Menge der Eigenwerte von A und  $\{\phi_i\}$  seien die entsprechenden ortonormierten Eigenvektoren.

Dann kann die Steifigkeit des Netzwerks ausgedrückt werden durch die inversen Eigenwerte:

$$1/\lambda_i^* =: \lambda_i$$

- 10 so dass  $\lambda_i$  die Steifigkeit der Systemantwort in Richtung des i-ten Eigenvektors unter einer Störung oder Stress beschreibt.

Mit Gleichung (I) kann  $x$  dargestellt werden durch Projektionen auf die Eigenvektoren von A gemäß Gleichung (S2):

$$x = \sum_k \frac{1}{\lambda_k} \phi_k \langle u, \phi_k \rangle \quad (S2)$$

worin  $\langle u, \phi_k \rangle$  das Skalarprodukt zwischen zwei Vektoren ist.

- 15 Weiterhin sei  $\{\omega_k\}$ , eine Störung des Systems mit der Struktur eines weißen Rauschens um den stationären Zustand, worin  $k$  der Index der Datensätze ist und die Dimension von  $(\omega_k) = n$  ist.

Dann sei ohne Beschränkung der Allgemeinheit:

$$\langle \omega \rangle = 1$$

$$\omega_k \text{ und } \omega_l \text{ sind unkorreliert: } \langle \langle \omega_k, \omega_l \rangle \rangle_{\text{Datensätze}} = \delta_{k,l}$$

- 20 wobei  $\omega_i$  die Bedeutung einer Störung des Systems in Richtung des i-ten Eigenvektors hat, wobei die Störungseinflüsse in Richtung der Eigenvektoren des Systems unkorreliert sein sind, so dass der Ausdruck  $\langle \langle \omega_k, \omega_l \rangle \rangle > 0$  ist, wenn  $k$  gleich  $l$  ist.

- 25 Unter diesen Annahmen gehorcht die Auslenkung  $\eta_i^k$  als Projektion dem Zustand auf den i-ten Eigenvektor von A der durch das Rauschen induzierten Störung von  $x$  durch  $\omega_k$ , entsprechend der mittleren Amplitude der rauschinduzierten Auslenkung des Systems in Richtung des i-ten Eigenvektors, den im folgenden dargestellten Bedingungen.

Entsprechend den Annahmen der Thermodynamik verteilt sich die durch ein weißes Rauschen induzierte Dehnungsenergie in einem elastischen Netzwerk gleichmäßig auf alle Eigenvektoren, so dass für die Erwartungswerte der Momente der Amplituden die folgende Gleichung (S3) gilt:

$$\begin{aligned} \langle \eta_i \rangle_T &= 0 \\ \langle |\eta_i|^2 \rangle_T &= \frac{|\omega|^2}{Z} \int |\eta_i|^2 \exp\left(-\frac{1}{2}\lambda_i |\eta_i|^2\right) d|\eta_i| = \mu \frac{|\omega|^2}{\lambda_i} \end{aligned} \quad (S3)$$

5 mit Z als der Zustandssumme gemäß der folgenden Gleichung (S4):

$$Z = \int \exp\left(-\frac{1}{2}\lambda_i |\eta_i|^2\right) d|\eta_i| \quad (S4)$$

und  $\langle |\eta_i|^2 \rangle_T$  als Mittelwert über alle Datensätze erhältlich aus den bereitgestellten Systemen, beispielsweise einer Anzahl bereitgestellter Gewebe.

Aus diesen Gleichungen für die Amplitudenverteilung (S3) kann die Statistik für die rausch-  
 10 induzierten Auslenkungen  $\xi_i$  in den Originalkoordinaten um den stationären Zustand durch Projektion der Amplitudenstatistik auf die Eigenvektoren gemäß der folgenden Gleichung (S5) berechnet werden:

$$\begin{aligned} \langle \xi_i^2 \rangle_T &= \mu \sum_k \frac{1}{\lambda_k} (\phi_k^i)^2 \\ \langle \xi_i, \xi_j \rangle_T &= \mu \sum_k \frac{1}{\lambda_k} \phi_k^i \phi_k^j \end{aligned} \quad (S5)$$

mit  $\phi_k^i$  als der i-ten Komponente des k-ten Eigenvektors. Hier bedeutet  $\langle \xi_i, \xi_j \rangle_T$  wieder den  
 15 Mittelwert, gebildet über alle Datensätze, erhältlich aus den bereitgestellten Systemen, beispielsweise einer Anzahl bereitgestellter Gewebe.

Man erhält eine Beziehung gemäß der nachstehenden Gleichung (S6):

$$\langle \xi_i, \xi_j \rangle_T = |\xi_i| |\xi_j| \text{cor}_T(\xi_i, \xi_j) \quad (S6)$$

mit  $\text{cor}_T(\xi_i, \xi_j)$  als den Korrelationskoeffizienten von  $\xi_i$  und  $\xi_j$  auf den Datensätzen für die  
 20 Komponenten i und j, und

$$|\xi_i| = \left( \langle \xi_i^2 \rangle_T \right)^{1/2} = \sigma_T(\xi_i) =: \sigma_i$$

als Länge des Vektors  $\xi_i$  auf dem Datensatz der Komponente  $i$ .

Eine Projektion des Stressvektors  $u = \{u_1, \dots, u_n\}$  auf die Eigenvektoren von  $A$ :

$$u = \sum_k \omega_k \varphi_k$$

$$\omega_j = \sum_i u_i \varphi_j^i$$

und ein Einsetzen in die Gleichung (S2) und Vertauschen der Summation ergibt für die  
 5 Auslenkung von  $x_i$ , induziert durch eine externe Störung oder Stress die nachstehende Gleichung (S7):

$$x_i = \sum_k \frac{1}{\lambda_k} \varphi_k^i \langle u, \varphi_k \rangle = \sum_k \frac{1}{\lambda_k} \varphi_k^i \sum_j u_j \varphi_k^j = \sum_j u_j \sum_k \frac{1}{\lambda_k} \varphi_k^i \varphi_k^j \quad (S7).$$

Ein Einsetzen von Gleichung (S5) in Gleichung (S7) und Benutzen der Korrelation der rausch-  
 induzierten Auslenkungen um den stationären Zustand, dargestellt durch Gleichung (S6), führt zu  
 10 nachstehender Gleichung (S8):

$$x_i = \sum_j \frac{1}{\mu} u_j \langle \xi_i, \xi_j \rangle_T = \frac{1}{\mu} |\xi_i| \left| \sum_j u_j \right| |\xi_j| \text{cor}_T(\xi_i, \xi_j) \quad (S8).$$

Sei nun formal:

$$\xi_u =: \sum_j u_j \xi_j$$

die gewichtete Summe über die  $\xi_j$ , wobei die Gewichte  $u_j$  die Störungskomponenten der  $j$ -ten  
 15 Komponente des Systems ist.  $\xi_u$  ist ein Vektor mit einer Länge die gleich der Anzahl der bereit-  
 gestellten Systemen, beispielsweise Gewebeproben, und beschreibt die effektive Störung oder den  
 effektiven Stress auf jedes System, beispielsweise eine Gewebeprobe, und hängt nicht von den  
 Komponenten  $i$  ab.

Unter Verwendung von  $\xi_u$  wird die Analyse vereinfacht zu nachstehender Gleichung (S9):

$$20 \quad x_i = \frac{1}{\mu} |\xi_i| \left| \sum_j u_j \right| |\xi_j| \text{cor}_T(\xi_i, \xi_j) = \frac{1}{\mu} |\xi_i| |\xi_u| \text{cor}_T(\xi_i, \xi_u) \quad (S9).$$

Hieraus folgt wegen  $|\xi_i| = \sigma_i$  die folgende Proportionalitätsrelation (S10):

$$\frac{x_i}{\sigma_i} \sim \text{cor}_T(\xi_i, \xi_u) \quad (\text{S10})$$

mit dem "effektiven Stressvektor"  $\xi_u$ , der unabhängig von den Komponenten  $i$  ist und aus den Daten der Aktivität der Komponenten identifiziert werden muss.

Die Proportionalitätskonstante der Gleichung (S10) entsprechend dem Term  $|u| \sigma_u$  der Gleichungen (IV) bis (VI) und ein Wert  $\xi_u^j$  für jeden Datensatz  $j$  ist mittels Auflösen eines linearen Gleichungssystems berechenbar.

Die Berechnung kann vorzugsweise im Rahmen einer Parameterschätzung durchgeführt werden. Bestimmbar sind die Daten der Aktivität der Komponenten, beispielsweise die Expressionswerte für alle Gene in dem System, beispielsweise einem Gewebe bzw. einer Probe des untersuchten Gewebes, an den stationären Zuständen. Daher ist die Anzahl der verfügbaren Datensätze für die Parameterschätzung gleich der Zahl der Komponenten mal der Zahl der Gewebeprobe-  
10 um die Zahl der Gene mal größer als der Mindestbedarf der benötigten Datensätze.

Da die Parameterschätzung letztlich auf Auflösen eines kleinen linearen Gleichungssystems reduziert werden kann, kann vorteilhafter Weise eine deutlich höhere Stabilität als bei einer direkten Schätzung aller Komponenten der Matrix  $A$  erwartet werden.  
15

Die Änderung der Aktivität einer Komponente  $i$  kann in Form der Korrelationskoeffizienten von Komponentenpaaren und der jeweiligen Standardabweichung gemäß der folgenden Gleichung (II) ausgedrückt werden.

$$x_i = \sigma_i \sum_j u_j \sigma_j \text{cor}(\xi_i, \xi_j) \quad (\text{II})$$

20 worin:

$x_i$ : ist die Verschiebung der Aktivität der  $i$ -ten Komponente als Reaktion auf die Störung,

$\sigma_i$ : ist die Standardabweichung der Komponente  $i$  in einem "stratifizierten" System,

$\text{cor}(\xi_i, \xi_j)$ : ist der lineare Korrelationskoeffizient zwischen den Änderungen der Aktivität der Komponenten  $i$  und  $j$  in dem stratifizierten System,  
25

$u_j$ : ist die Störung, die auf die Komponente  $j$  wirkt.

Der Begriff "stratifizieren" hat im Sinne der Berechnungen des erfindungsgemäßen Verfahrens die Bedeutung, dass für jede Komponente der Mittelwert der Aktivität vor und nach der ausgeübten Störung berechnet wird. Danach wird für jede Komponente und jeden Wert der Aktivität der jeweilige Mittelwert subtrahiert. In bevorzugten Ausführungsformen des Verfahrens hat der

5 Begriff "stratifizieren" im Sinne der Berechnungen des erfindungsgemäßen Verfahrens die Bedeutung, dass für jedes bestimmte Gen der Mittelwert der Expression für jeden applizierten pharmazeutischen Wirkstoff, oder gemittelt über eine applizierte Stoffgruppe umfassend mehrere gleichwirkende Wirkstoffe, berechnet. Danach wird für jedes Gen und jeden Expressionswert der

10 jeweilige Mittelwert subtrahiert. Hierdurch wird erreicht, dass nur noch die Schwankungen um den jeweils durch die Mittelwerte beschriebenen stationären Zustand oder Steady state berücksichtigt werden.

Unter Verwendung von  $|u| = (\sum u_k^2)^{1/2}$ , wobei k Koeffizienten für jede Komponente darstellen, die die Wirkung der Störung auf die Komponente darstellen, für die gesamte Störung ist die "effektive" Störung durch die folgende Gleichung (III)

$$15 \quad \xi_u = \frac{\sum_j u_j \xi_j}{|u|} \quad (\text{III})$$

worin:

$\xi_u$  ist der formale Vektor der Aktivitätsänderung an einer fiktiven Komponente, die den Angriffspunkt der Störung darstellt und über gewichtete Mittelung über die x – Werte der beteiligten Komponenten berechnet wird,

20  $\sum_j u_j \xi_j$  beschreibt die Berechnung des gewichteten Mittelwerts der Aktivitäten der Komponenten, die durch die Störung oder den Stress direkt beeinflusst werden,

$|u|$  die Intensität der Störung oder des Stress wiedergibt,

umformulierbar.

Der Term  $|u|$  ist hierbei identisch mit  $1/\mu$  in Gleichung (S9) der formalen Herleitung.

25 In bevorzugten Ausführungsformen des Verfahrens sind die Daten der Aktivitätsänderung an einer fiktiven Komponente, die den Angriffspunkt der Störung darstellt, Expressionswerte für die Genexpression.

Diese Umformulierung der Störung erlaubt die Summe der Einwirkung einer Komponente j auf die durch die Störung verursachte Änderung der Aktivität einer Komponente i durch die nachstehende Gleichung (IV):

$$x_i = |u| \sigma_i \sigma_u \text{cor}(\xi_i, \xi_u) \quad (\text{IV})$$

5 worin:

$x_i$ : ist die Verschiebung der Aktivität der i-ten Komponente als Reaktion auf die Störung oder den Stress,

$|u|$  ist die Intensität der Störung,

$\sigma_u$ : ist die Standardabweichung der durch das Rauschen u erzeugten Antwort,

10  $\sigma_i$ : ist die Standardabweichung der Komponente i,

$\text{cor}(\xi_i, \xi_u)$ : ist der lineare Korrelationskoeffizient zwischen den Änderungen der Aktivität der Komponenten i und j in dem stratifizierten System,

auszudrücken. Hierbei entspricht  $\sigma_u$  dem  $|\xi_u|$  in Gleichung (S9).

15 Äquivalent kann die Gleichung (IV) durch die folgende algebraische Gleichung (V) ausgedrückt werden:

$$\frac{x_i}{\sigma_i} = |u| \sigma_u \text{cor}(\xi_i, \xi_u) = r \text{cor}(\xi_i, \xi_u) \quad (\text{V})$$

worin:

$r$ : ist die Steigung.

20 Die Gleichungen (IV) und (V) beschreiben die Änderung der Aktivität der Komponenten durch eine reversible Störung, wobei die Berechnung erfolgt über die Stärke der Störung  $|u|$ , die Standardabweichung  $\sigma_i$  der  $\xi_i$  der Komponente i und einen Vektor  $\xi_u$  und  $\sigma_u$ , der die effektive Störung auf die Komponenten wiedergibt.

25 Die Gleichungen (IV) und (V) sind nicht länger abhängig von einer tatsächlichen Komponente i, so dass es für die Berechnung des Verhaltens des biologischen Mehrzwecks ausreichend ist, einen Vektor  $\sigma_u \xi_u$  zu bestimmen und eine Zahl für  $|u|$  als "effektive Stärke der Störungen" zu bestimmen. Diese Bestimmung ist möglich über die bestimmten Daten der Änderung der Aktivität der Kompo-

nenten des Netzwerks, wobei  $|u|$  für sich genommen nicht messbar ist und die Größe, die in das Modell eingeht,  $r = |u| \sigma_u$  ist, wobei  $r$  durch lineare Regression aus Gleichung (V) mit Hilfe der Messdaten bestimmt werden kann.

Somit erlaubt das zur Verfügung gestellte Verfahren das Verhalten eines biologischen Netzwerks auf eine reversible Störung mit Hilfe des bereitgestellten linearen Modells aus den bestimmten Daten der Änderung der Aktivität der Komponenten als Reaktion auf eine reversible Störung zu berechnen.

Der Anstieg  $r = |u| \sigma_u$  stellt ein Maß für die Sensitivität der Änderung der Aktivität der Komponenten zur Verfügung, mit einem Bezug auf den formalen Abstand der Komponente  $i$  auf den Ort des Wirkens des Stresses ausgedrückt durch den Korrelationskoeffizienten  $\text{cor}(\xi_i, \xi_u)$ . Unter Voraussetzung eines Netzwerks der Komponenten mit rein linearer Wechselwirkung der Komponenten untereinander und ohne Streuung sollte die Steigung  $r$  für alle Komponenten konstant sein.

Aus den Gleichungen (IV) bzw. (V) ergibt sich, dass der Vektor  $\xi_i$  für Komponenten mit hohen Werten des Parameters  $x_i / \sigma_i$  zu dem Vektor  $\xi_u$  hochkorreliert sein sollte. Der Vektor  $\xi_u$  ist die verbleibende, nicht aus der Bestimmung der Aktivitätsänderung der Komponenten messbare Größe. Obwohl  $\xi_u$  unbekannt ist ergibt sich, dass der Vektor  $\xi_i$  für Gruppen von Komponenten mit ähnlichen Werten von  $x_i / \sigma_i$  in einem "Winkel" um  $\xi_u$  orientiert ist, wobei der Kosinus des Kegelwinkels durch den Parameter  $\text{cor}(\xi_i, \xi_u)$  gegeben ist. Der Parameter  $\xi_u$  ist unbekannt, da der Vektor  $\xi_i$  der einzelnen Komponenten eine verschiedene Korrelation zu dem Vektor  $\xi_u$  aufweist.

Die Bestimmung der Änderung der Aktivität der Komponenten gibt die Änderung der Aktivität für jede Komponente  $i$  und somit den Parameter  $x_i$  an, sowie die Standardabweichung  $\sigma_i$  der Komponente  $i$ .

Die Standardabweichung  $\sigma_i$  wird bei der Erstellung des Modells aus mehreren Messungen bestimmt. Vorzugsweise werden hierzu wenigstens zwei biologische Systeme, vorzugsweise wenigstens drei, bevorzugt wenigstens vier biologische Systeme, vorzugsweise ausgewählt aus der Gruppe umfassend Zelle, Zellkultur, Gewebe, Organ und/oder Organismus bereitgestellt und das Verfahren insbesondere die Schritte (a) bis (g) in den bereitgestellten Systemen durchgeführt. Aus den erhaltenen Messdaten der Änderung der Aktivität der Komponenten, beispielsweise der Änderung der Genexpression, nach der verwendeten reversiblen Störung ist dann die Standardabweichung  $\sigma_i$  für die Komponente  $i$  berechenbar.

Von besonderem Vorteil ist hierbei, dass die Standardabweichung  $\sigma_i$  für die Komponente  $i$  anhand der verwendeten Störung in einem System bestimmt wird, und nachfolgend bei der Anwendung des Modells für anderer Störungen des Systems verwendbar ist.

Von weiterem Vorteil ist hierbei, dass die einmal bestimmte Standardabweichung  $\sigma_i$  für die Komponente  $i$  erlaubt, das erfindungsgemäße Verfahren für eine andere Störung der Komponente  $i$  in dem verwendeten System zu verwenden, ohne dass  $\sigma_i$  erneut zu bestimmen ist. Vorteilhafter Weise ist das Verhalten eines Netzwerks umfassend Komponenten bekannter Standardabweichung  $\sigma$  für die Komponenten aus den in Schritten (c) und (e) bestimmten Aktivität der biologischen oder biochemischen Komponenten des biologisches Netzwerks vor und nach Ausüben der reversiblen Störung bestimmbar.

Somit ergibt sich der Vektor  $\xi_i$  für alle Komponenten  $i$  und die Gleichung (V) erlaubt die Berechnung von  $\sigma_u$   $\xi_u$ . Diese Berechnung kann mittels Optimierungsverfahren durchgeführt werden. Geeignete Optimierungsverfahren sind beispielsweise alle Verfahren der Kombinatorischen Optimierung, beispielsweise ausgewählt aus der Gruppe umfassend Genetische Algorithmen und/oder das Simulierte Ausglühen oder "Simulated Annealing". Geeignete Genetische Algorithmen sind beispielsweise beschrieben in Ingo Rechenberg, Evolutionsstrategie '94, Frommann Holzboog, 1994.

Die Berechnung von  $\xi_u$  kann insbesondere unter der Voraussetzung, dass  $|u|$  wie auch  $\xi_u$  näherungsweise konstant in einem biologischen System sind, berechnet werden.

Die Rekonstruktion von  $\xi_u$  aus den Daten der bestimmten Änderung der Aktivität der Komponenten setzt voraus, dass die Gleichung (V) in ein überbestimmtes lineares Gleichungssystem überführt wird.

$\xi_u$  wird vorzugsweise durch kombinatorische Optimierung bestimmt, wobei ein bevorzugter Algorithmus der sogenannte genetische Algorithmus ist. Dieser ist beispielsweise beschrieben in Ingo Rechenberg, Evolutionsstrategie '94, Frommann Holzboog, 1994. Weitere geeignete Optimierungsverfahren, die die Berechnung von  $\xi_u$  aus den bestimmten Daten der Änderung der Aktivität der Komponenten erlauben, sind beispielsweise ausgewählt aus der Gruppe umfassend Simulierte Ausglühen oder "Simulated Annealing" genannt und/oder Sintflut-Algorithmus oder "Grand Deluge" genannt.

Vorzugsweise wird  $\xi_u$  in Form einer linearen Kombination aus den bestimmten Daten der Änderung der Aktivität der Komponenten für eine gewählte Anzahl Komponenten bestimmt. Die

Anzahl der Komponenten, die für eine solche Bestimmung benutzt werden, kann vorzugsweise im Bereich von 1 bis 4.000 Komponenten, bevorzugt im Bereich von 5 bis 100 Komponenten liegen.

Aus der Anzahl Komponenten kann eine geeignete Untergruppe an Komponenten, beispielsweise  $S_u$  benannt, beispielsweise mit einer Anzahl Komponenten im Bereich von  $\geq 10$  Komponenten bis  $\leq 4000$  Komponenten, vorzugsweise im Bereich von  $\geq 20$  Komponenten bis  $\leq 200$  Komponenten verwendet werden, um die statistische Gewichtung  $w_i$  für eine lineare Kombination gemäß der folgenden Gleichung (VI):

$$\xi_u' = \sum_{i \in S_u} w_i \xi_i \quad (\text{VI})$$

worin:

10  $\xi_u'$  ist der optimierte formale Vektor des biologischen Rauschens an einer fiktiven Komponente, die den Angriffspunkt der Störung darstellt,

$w_i$  ist die statistische Gewichtung der Komponenten,

15  $\xi_i$  ist der Vektor der Verschiebung der i-ten Komponente als Reaktion auf das Rauschen um den Mittelwert der Aktivität der Komponente i, beispielsweise der Expression des Gens i, im stratifizierten System,

zu berechnen. Die berechnete Gewichtung  $w_i$  erlaubt die Berechnung der linearen Korrelationskoeffizienten der Gleichung (V) wie auch die der weiteren Parameter der Gleichung. Die erhaltenen Werte können dazu benutzt werden den genetischen Algorithmus und eine optimale Anzahl Komponente zur Optimierung von  $\xi_u$  zu bestimmen. Diese Optimierung ist vorzugsweise  
20 Teil der verwendbaren Optimierungsverfahren.

Unter Verwendung des optimierten  $\xi_u'$  ist die Gleichung (V) bzw. (IV) für alle Komponenten berechenbar.

Das erfindungsgemäße Verfahren erlaubt somit das Verhalten eines biologischen Netzwerks zu berechnen anhand von experimentell erhältlichen Daten der Änderung der Aktivität der einzelnen  
25 Komponenten des Netzwerks. Von besonderem Vorteil ist hierbei, dass eine solche Berechnung auch bei einer sehr großen Anzahl an Komponenten anhand des bereitgestellten linearen Modells zur Beschreibung des Verhaltens des Netzwerks ermöglicht ist, wobei unter Berücksichtigung der Biodiversität der Reaktion der Komponenten eine Berechnung ermöglicht wird, ohne dass inner-

halb des bereitgestellten linearen Modells eine Matrix, die die Parameter enthält, die die Reaktion der Komponenten auf eine Störung beschreiben, explizit berechnet werden muss.

In bevorzugten Ausführungsformen des erfindungsgemäßen Verfahrens ist die Biodiversität eine biologische Variation ausgewählt aus der Gruppe umfassend natürliche Variation einer Aktivität  
5 einer Komponente oder eines Netzwerks, eine natürliche Variation eines biologischen Systems und/oder eine Variation der biologischen Reaktion eines Systems auf Umgebungsfaktoren, die es erlaubt, das bereitgestellte lineare Modell mit Hilfe der durch die Biodiversität erzeugten Variationen ohne systematische Experimente zu bestimmen.

Dies stellt einen besonderen Vorteil des erfindungsgemäßen Verfahrens zur Verfügung mit dem  
10 das Verhalten eines Netzwerks vieler Komponenten beispielsweise einer großen Anzahl von Genen, wie sie beispielsweise als Reaktion auf einen toxischen Stress reguliert werden können, zu bestimmen, ohne dass systematische Experimente durchgeführt werden müssen.

Insbesondere ermöglicht das erfindungsgemäße Verfahren, dass durch Bereitstellen eines biologischen Systems, Ausüben einer Störung auf dieses System, und einer einmaligen Bestimmung der  
15 Änderung der Aktivität der Komponenten auf die reversible Störung das Verhalten anhand des bereitgestellten linearen Modells beschrieben werden kann.

Eine Störung kann beispielsweise ein Stress sein, der auf das System wirkt. Vorzugsweise ist die Störung ein externer Stress, bevorzugt ausgewählt aus der Gruppe umfassend toxischer Stress, vorzugsweise ausgewählt aus der Gruppe umfassend Stress durch nicht-genotoxische oder genotoxische Hepatokarzinogene, Wärmestress, Stress durch Hunger, Stress durch Applikation eines  
20 pharmazeutischen Wirkstoffs, einer Chemikalie und/oder eines Medikaments.

Bevorzugte biologische Systeme sind ausgewählt aus der Gruppe umfassend Zelle(n), Gewebe, Organ(e) und/oder Organismus, wobei bevorzugte Gewebe oder Organe solche sind, die biologische und/oder biochemische Komponenten enthalten. Bevorzugte Gewebe oder Organe sind  
25 beispielsweise ausgewählt aus der Gruppe umfassend Gehirn und/oder Leber. Es versteht sich, dass jedes biologische System im Rahmen der vorliegenden Erfindung verwendet werden kann, beispielsweise prokariotische wie eukariotische Zellen oder Organismen. Ein biologisches System kann beispielsweise eine Zellkultur sein oder ein Säugetierorganismus wie Maus oder Ratte, die durch geeignete experimentelle Ausführung einer reversiblen Störung ausgesetzt werden können.

30 Bevorzugte biologische Komponenten sind Gene. Insbesondere die Untersuchung der Genexpression ist Gegenstand weitläufiger Untersuchungen zur Reaktion von biologischen Systemen auf

eine Störung bzw. Stress. Bevorzugte biochemische Komponenten sind ausgewählt aus der Gruppe umfassend RNA, DNA, Metaboliten und/oder Proteine.

Biologische und/oder biochemische Komponenten können auf eine reversible Störung reagieren, indem sie ihre Aktivität ändern. Abhängig von der Art des Stresses und der hierdurch beeinflussten  
5 Komponenten und/oder der Stärke der ausgeübten Störung werden unterschiedliche biologische und/oder biochemische Komponenten durch eine solche Störung beeinflusst. Durch eine solche Störung können abhängig von der Art und des Ausmaßes der Störung viele oder wenige Komponenten eines Netzwerks beeinflusst werden. Die Anzahl der Komponenten, die durch eine Störung  
10 direkt betroffen werden kann in weiten Bereichen variieren, beispielsweise in einem Bereich von  $\geq 1$  Komponente bis alle Komponenten, entsprechend  $\leq 100$  % der Komponenten, bezogen auf 100 % Komponenten, bevorzugt im Bereich bis  $\leq 20$  % der Komponenten, bevorzugter im Bereich bis  $\leq 10$  % der Komponenten, vorzugsweise im Bereich bis 5 % der Komponenten, auch bevorzugt im Bereich bis  $\leq 3$  % der Komponenten, noch bevorzugt im Bereich bis  $\leq 2$  % der Komponenten, bezogen auf 100 % der Komponenten, liegen.

15 In weiterhin bevorzugten Ausführungsformen des erfindungsgemäßen Verfahrens kann eine Störung auf die Änderung der Aktivität aller Komponenten berechnet werden, solange ihre Aktivität vorzugsweise ihre Expression hinreichend genau gemessen werden kann. Die hinreichend genau bestimmbare Anzahl der Komponenten beispielsweise bei Genexpressionsnetzwerken liegt im Bereich bis 40 % der Komponenten, vorzugsweise im Bereich bis 30 % der Komponenten,  
20 bezogen auf 100 % der Komponenten. Es ist ein besonderer Vorteil des erfindungsgemäßen Verfahrens, dass die grobe Berechnung des Verhaltens eines Netzwerks auch dann noch ermöglicht wird, wenn mehr als 30 % der Komponenten eines Netzwerks, insbesondere wenn mehr als 40 % der Komponenten eines Netzwerks von der reversiblen Störung betroffen sind.

Die Aktivität der biologischen oder biochemischen Komponenten des Netzwerks kann hierbei in  
25 Abhängigkeit von der reversiblen Störung ebenfalls in einem variierenden Ausmaß beeinflusst werden. In bevorzugten Ausführungsformen des erfindungsgemäßen Verfahrens ist die Aktivität der Komponenten in einem Bereich von 0,1 % bis 30 %, bevorzugt 0,5 % bis 25 %, vorzugsweise 1 % bis 20 %, noch bevorzugter 5 % bis 15 % bezogen auf die Aktivität der biologischen oder biochemischen Komponenten im Grundzustand, d.h. in einem Zustand bevor oder das keine Störung  
30 auf das System ausgeübt wird, beeinflusst.

Das erfindungsgemäße Verfahren ist in bevorzugten Ausführungsformen ein Verfahren in dem Gebiet der quantitativen Toxikogenomik. In bevorzugten Ausführungsformen sind die biochemischen oder biologischen Komponenten entsprechend Gene und RNA- und/oder DNA-Moleküle. Im Rahmen der vorliegenden Erfindung bedeutet Änderung der Aktivität eines Gens

vorzugsweise, dass ein solches Gen in seiner Expression hoch- oder runter-reguliert wird. Die Expressionsrate eines Gens ist vorzugsweise als Gehalt an RNA oder dem entsprechenden Genprodukt bestimmbar. In besonders bevorzugten Ausführungsformen wird der in dem entsprechenden System, vorzugsweise einer Zellkultur oder Zellen eines Gewebes vorliegende RNA-Gehalt bestimmt.

Bevorzugt wird die Änderung der Aktivität mindestens einer biologischen oder biochemischen Komponente entsprechend mittels Verfahren bestimmt, die Auskunft über die in einem System vorliegenden RNA- oder DNA-Gehalt geben können, vorzugsweise aus der Gruppe umfassend semiquantitative RT-PCR, Northern Hybridisierung, Differential display, substraktive Hybridisierung, subtrahierte Bibliotheken, cDNA Arrays und/oder Oligo-Arrays.

In anderen bevorzugten Ausführungsformen des erfindungsgemäßen Verfahrens kann die biochemische Komponente ein Protein sein, oder ein Metabolit einer Wirksubstanz, die als Störung verabreicht wurde.

Es kann entsprechend weiterhin bevorzugt sein, dass man die Änderung der Aktivität einer Komponente mittels Verfahren bestimmt, die ausgewählt sind aus der Gruppe umfassend Verfahren, die zur Bestimmung eines Proteingehalts eines Systems verwendbar sind, vorzugsweise ausgewählt aus der Gruppe umfassend Western Hybridisierung, ELISA-Technik (Enzyme Linked Immuno Sorbent Assay) und/oder spektroskopische Methoden, beispielsweise HPLC (High Pressure Liquid Chromatography), fluoreszenzbasierte absorptive oder massenspektrometrische Nachweise.

In bevorzugten Ausführungsformen des erfindungsgemäßen Verfahrens kann man zwischen der gemäß Schritt (f) bestimmten Änderung der Aktivität der einzelnen Komponenten und dem gemäß Schritt (g) berechneten Verhalten des biologischen Netzwerks anhand des bereitgestellten linearen Modells vergleichen, wobei man erwartet, dass eine Übereinstimmung des berechneten Verhaltens mit den in Schritt (f) bestimmten Änderungen der Aktivität der biologischen oder biochemischen Komponenten besteht. Ergibt ein solcher Vergleich, dass eine Übereinstimmung zwischen der bestimmten Änderung der Aktivität einer Komponente und der entsprechenden Berechnung durch das bereitgestellte Modell besteht, entsprechend also eine Übereinstimmung vorzugsweise experimentell bestimmter Daten und der Berechnung des Modells besteht, unterliegt die experimentell bestimmte Reaktion der Komponente auf die Störung den Vorhersagen des Modells.

In anderen Ausführungsformen der erfindungsgemäßen Verfahren kann bei einem solchen Vergleich entsprechend Schritt (h) des Verfahrens feststellbar sein, dass eine statistisch signifikante Abweichung einer oder mehrerer Komponente(n) der gemäß Schritt (f) bestimmten

Änderung der Aktivität und dem gemäß Schritt (g) berechneten Verhalten der Komponente(n) im Netzwerk vorliegt, die anzeigt, dass diese Komponente(n) nicht dem bereitgestellten linearen Modell unterliegt. Eine solche Komponente, die dem bereitgestellten linearen Modell nicht unterliegt, kann ein Indikator für einen störungsinduzierten Übergang in einen neuen Zustand der Komponente sein und einen solchen Übergang anzeigen. Eine solche Abweichung von dem durch das bereitgestellte lineare Modell berechneten Verhalten, kann insbesondere bedeuten, dass die Störung für die Komponente nicht reversibel ist. Bei einer nicht reversiblen Störung kehrt das System nach Entfernen des Stresses nicht in seinen Ausgangszustand zurück, und/oder eine einzelne Komponente kehrt nach Entfernen der Störung nicht in den Ausgangszustand der Aktivität vor der reversiblen Störung zurück. Eine solche Komponente kann als Indikator dafür dienen, dass das System in einen anderen Zustand des biologischen Systems übergegangen ist, beispielsweise in einen Zustand, der einer Erkrankung verursacht durch die Störung entspricht.

Es ist ein Vorteil des erfindungsgemäßen Verfahrens, dass eine feststellbare statistisch signifikante Abweichung einer oder mehrerer Komponenten eine Aussage darüber ermöglicht, ob das System eine oder mehrere Komponenten aufweist, die anzeigen können, dass das System nach der ausgeübten Störung nicht reversibel reagiert, sondern einen davon abweichenden Zustand, vorzugsweise einen Zustand, der eine Erkrankung des Systems charakterisiert, einnimmt.

Die bevorzugten Ausführungsformen des erfindungsgemäßen Verfahrens bestimmt man die statistische Signifikanz mittels eines Signifikanztests vorzugsweise ausgewählt aus der Gruppe umfassend T-Test, Z-Test, und/oder chi-Quadrat-Test.

In weiteren Ausführungsformen des Verfahrens kann man in einem weiteren Schritt erhalten, dass eine statistisch signifikante Regulation der Aktivität einer oder mehrerer Komponente(n) gemäß der in Schritt (f) bestimmten Änderung der Aktivität und dem gemäß Schritt (g) berechneten Verhalten der Komponente im Netzwerk vorliegt.

Der Abstand von einem direkten Angriffspunkt der Störung ist erhältlich durch den Korrelationskoeffizienten  $\text{cor}(\xi_i, \xi_u)$ . Je größer der Absolutbetrag ist, desto dichter ist die Komponente am Angriffspunkt.

Eine solche statistisch signifikante Isolation der Aktivität einer oder mehrerer Komponenten kann bedeuten, dass diese Komponente dicht am mechanistischen Angriffspunkt der Störung liegt. Eine solche Komponente, die signifikant stärker durch die ausgeübte Störung in ihrer Aktivität reguliert wird, weist eine hohe Sensitivität gegenüber der Störung auf. Eine solche signifikant regulierte Komponente kann eine Komponente, beispielsweise ein Gen, sein, die mit einem entsprechenden

Berechnungsverfahren zur Berechnung nicht direkt beobachtbare Größe, beispielsweise physiologische Änderungen eines Organismus, einen Biomarker ausbilden.

In einer weiteren bevorzugten Ausführungsform des Verfahrens kann dieses zur Bestimmung von Biomarkern dienen.

- 5 In einer weiteren bevorzugten Ausführungsform des Verfahrens können die Schritte (a) bis (h) für wenigstens zwei reversible Störungen und optional wenigstens zwei Systeme wiederholt werden und man erhält in einem weiteren Schritt des Vergleichs, dass eine statistisch signifikante Regulation der Aktivität einer oder mehrerer Komponente(n) gemäß der in Schritt (f) bestimmten Änderung der Aktivität und dem gemäß Schritt (g) berechneten Verhalten der Komponente in
- 10 Bezug auf verschiedene Arten von Störungen vorliegt, die eine Klassifikation der Störung anhand des Auftretens der statistisch signifikanten Regulation der Komponente(n) erlaubt.

Vorzugsweise ist feststellbar, dass wenigstens eine der bestimmten Komponenten in Bezug auf eine bestimmte Art von Störung einer statistisch signifikante Regulation aufweist, und in Bezug auf weitere Arten von Störungen davon statistisch signifikant verschiedene Regulationen aufweist,

15 so dass eine statistisch signifikante charakteristische Reaktion auf eine bestimmte Störung feststellbar ist. Eine solche statistisch signifikante Regulation wenigstens einer Komponente auf eine bestimmte Störung erlaubt es, die Störung anhand des Auftretens einer solchen als Biomarker bezeichneten Komponente zu klassifizieren. Das Erhalten eines solchen Biomarkers kann in bevorzugten Ausführungsformen des Verfahrens unter Bestimmung der Änderung der Aktivität

20 wenigstens einer Komponente und der Berechnung des Verhaltens des Netzwerks, dem diese Komponente angehört, entsprechend dem zur Verfügung gestellten linearen Modell zur Verfügung gestellt werden.

In bevorzugten Ausführungsformen erhält man eine statistisch signifikante Regulation der Aktivität mehrerer Komponenten, wobei eine solche Regulation eine positive oder negative

25 Regulation, beispielsweise in Bezug auf die Expressionsrate von Genen eine Auf- oder Abregulation der Genexpression, sein kann. Die statistisch signifikante Regulation mehrerer Komponenten ist nicht notwendigerweise gleichgerichtet, sondern kann vorzugsweise eher einem charakteristischen Muster der Regulation der verschiedenen Komponenten entsprechen.

Vorteilhafterweise ermöglicht das erfindungsgemäße Verfahren in bevorzugten Ausführungs-

30 formen, dass eine große Anzahl von Komponenten durch das Modell berechenbar ist. In weiterhin vorteilhaften Ausführungsformen des Verfahrens ermöglicht das Verfahren weiterhin, dass die Berechnung auf möglichst wenige Komponenten beschränkbar ist. Dies ermöglicht das erfindungsgemäße Verfahren vorteilhafterweise dadurch, dass man eine statistisch signifikante Regulation

der Aktivität einer oder mehrerer Komponenten und der berechneten Änderung des Verhaltens des Netzwerks ermöglicht, dass die signifikant geregelten Komponenten über ihre signifikante Regulation durch eine bestimmte Störung erlauben, diese Störung beispielsweise in weiteren oder wiederholten Verfahren zu klassifizieren.

- 5 Das erfindungsgemäße Verfahren ist in bevorzugten Ausführungsformen ein Verfahren auf dem Gebiet der quantitativen Toxikogenomik. In bevorzugten Ausführungsformen des Verfahrens sind die Komponenten Gene und es wird die Genexpression vorzugsweise von Stressgenen bestimmt. Vorzugsweise ist das System ein Säugetier, beispielsweise eine Ratte oder Maus, das verschiedene Gewebe beispielsweise ausgewählt aus der Gruppe umfassend Leber und Gehirn umfasst, oder  
10 eine Zellkultur. Vorzugsweise wird eine externe Störung ausgeübt, indem ein reversibler toxischer Stress auf das System ausgeübt wird. Bevorzugt wird wenigstens ein, vorzugsweise mehrere pharmazeutischer Wirkstoff, vorzugsweise wenigstens ein Karzinogen appliziert. Es können in mehreren bereitgestellten Systemen mehrere pharmazeutische Wirkstoffe oder andere Chemikalien, vorzugsweise Karzinogene, bevorzugt ausgewählt aus der Gruppe umfassend Wirkstoffe, die  
15 einen nicht-genotoxischen Stress, genotoxischer Stress und/oder hepatotoxischer Stress ausüben, appliziert werden.

In einer besonders bevorzugten Ausführungsform des Verfahrens betrifft das Verfahren die Bestimmung der Änderung der Genexpression in einem Gewebe nach einem reversiblen toxischen Stress umfassend die folgenden Schritte:

- 20 (a) Bereitstellen eines Organismus, der ein Gewebe enthält, das ein biologisches Netzwerk umfassend eine Vielzahl von Genen umfasst;
- (b) Bereitstellen eines linearen Modells zur Beschreibung der Änderung der Genexpression des Netzwerks;
- (c) Bestimmen der Grundexpression der Gene;
- 25 (d) Ausüben eines toxischen Stresses, bevorzugt Applikation eines pharmazeutischen Wirkstoffs, vorzugsweise eines Karzinogens, wobei eine Änderung der Genexpression erzeugt wird;
- (e) Bestimmen der Genexpression nach Applikation des toxischen Stresses, bevorzugt des pharmazeutischen Wirkstoffs, vorzugsweise des Karzinogens, sobald die Gene des Netz-  
30 werks die Reaktion auf den Stress vollzogen haben;

- (f) Bestimmen der Änderung der Expression wenigstens eines Gens nach Ausüben des toxischen Stresses, bevorzugt Applikation des pharmazeutischen Wirkstoffs, vorzugsweise des Karzinogens;
- (g) Berechnen der Änderung der Genexpression aller Gene des Netzwerks anhand des bereitgestellten linearen Modells zur Beschreibung des Verhaltens des biologischen Netzwerks aus der bestimmten Änderung der Expression wenigstens eines Gens unter Berücksichtigung der Biodiversität der Änderung der Genexpression; und
- (h) optional Vergleichen der gemäß Schritt (f) bestimmten Änderung der Expression wenigstens eines Gens und der gemäß Schritt (g) berechneten Änderung der Genexpression der Gene des Netzwerks anhand des bereitgestellten linearen Modells, wobei man erwartet, dass eine Übereinstimmung berechneten Änderung der Genexpression mit den in Schritt (f) bestimmten Änderung der Expression wenigstens eines Gens besteht.

In bevorzugten Ausführungsformen des Verfahrens ist das Karzinogen ausgewählt aus der Gruppe umfassend nicht-genotoxische, genotoxische und/oder hepatotoxische Karzinome.

- 15 In bevorzugten Ausführungsformen des Verfahrens ist vorgesehen, dass man die Expression einer Anzahl Gene im Bereich von  $\geq 1$  Gen bis  $\leq 25.000$  Gene, vorzugsweise im Bereich von  $\geq 1$  Gene bis  $\leq 15.000$  Gene, bevorzugt im Bereich von  $\geq 1$  Gene bis  $\leq 5000$  Gene, besonders bevorzugt im Bereich von  $\geq 2$  Gene bis  $\leq 1000$  Gene, weiter vorzugsweise im Bereich von  $\geq 5$  Gene bis  $\leq 400$  Gene, noch vorzugsweise im Bereich von  $\geq 5$  Gene bis  $\leq 200$  Gene bestimmt.
- 20 Ein weiterer Gegenstand der vorliegenden Erfindung betrifft ein Computerprogrammprodukt mit computerlesbaren Programmmitteln zur Durchführung eines oder mehrerer Schritte des Verfahrens, wenn das Programm auf einem Computer ausgeführt wird. Die Erfindung lässt sich vorteilhafterweise in einem oder mehreren Computerprogrammen zur Ausführung in einem Computersystem mit Softwarekomponenten zur Durchführung eines oder mehrerer Schritte des Verfahrens durchführen, wenn das Programm auf einem Computer ausgeführt wird.
- 25 Ein weiterer Gegenstand der vorliegenden Erfindung betrifft somit ein Computerprogramm zur Ausführung in einem Computersystem mit Softwarekomponenten zur Durchführung eines oder mehrerer Schritte des Verfahrens, wenn das Programm auf einem Computer ausgeführt wird. Ein weiterer Gegenstand des Verfahrens betrifft ein Computersystem mit Mitteln zur Ausführung eines oder mehrerer Schritte des erfindungsgemäßen Verfahrens.
- 30

Wenn nicht anders ausgeführt, weisen die verwendeten technischen und wissenschaftlichen Ausdrücke die Bedeutung auf, wie sie gemeinhin von einem Durchschnittsfachmann in dem Gebiet, zu dem diese Erfindung gehört, verstanden wird.

Alle Veröffentlichungen, Patentanmeldungen, Patente und weiteren hier angegebenen Literaturangaben sind voll inhaltlich durch Bezugnahme aufgenommen.

Beispiele, die der Veranschaulichung der vorliegenden Erfindung dienen, sind nachstehend angegeben.

Berechnungen und Datenanalysen wurden unter Verwendung von Matlab, Mathworks, Waltham, USA, ausgeführt, wenn nicht anders angegeben.

**Beispiel 1**

Bestimmung der Genexpression in Rattenleber nach einem reversiblen toxischen Stress

Versuchsablauf, Behandlungsbedingungen und Probenaufbereitung erfolgte wie in "Ellinger-Ziegelbauer et al., Mutation Research 575, 2005 S. 61-84" beschrieben, sofern im folgenden nicht  
5 anders angegeben.

Für die in vivo Studien wurden männliche Wistar-Hanover-Ratten (CrI:WI[G1/BRL/Han]IGS BR, Charles River Laboratories Inc, Raleigh, USA) in Versuchsgruppen von jeweils 5 Tieren eingeteilt und erhielten einmal täglich für einen Zeitraum von 1, 3, 7 oder 14 Tagen per Magensonde ("Gavage") jeweils eine der folgenden Substanzen in der angegebenen Konzentration. Verwendet  
10 wurden fünf genotoxische Karzinogene: 2-Nitrofluoren (Sigma, St. Louis, USA), in einer Konzentration von 4 mg/kg/Tag für 3 und 7 Tage, Dimethylnitrosamin (Sigma, St. Louis, USA), in einer Konzentration von 4 mg/kg/Tag für 3 und 7 Tage, Aflatoxin B1 (Sigma, St. Louis, USA), in einer Konzentration von 0,24 mg/kg/Tag für 3 und 7 Tage, N-nitrosomorpholin (TCI America, Portland, USA), in einer Konzentration von 3,5 mg/kg/Tag für 3 und 7 Tage, und CI Direct Black (TCI  
15 America, Portland, USA), 146 mg/kg/Tag für 3 und 7 Tage; fünf nicht-genotoxische Karzinogene: Methapyrilene HCl (Sigma, St. Louis, USA), in einer Konzentration von 60 mg/kg/Tag für 3 und 7 Tage, Thioacetamid (Sigma, St. Louis, USA), in einer Konzentration von 19,2 mg/kg/Tag für 3 und 7 Tage, Diethylstilbestrol (Sigma, St. Louis, USA), in einer Konzentration von 10 mg/kg/Tag für 1 und 3 Tage, Wy 14643 (TCI America, Portland, USA), in einer Konzentration von  
20 60 mg/kg/Tag für 1 und 3 Tage, und Piperonylbutoxid (Sigma, St. Louis, USA), in einer Konzentration von 1200 mg/kg/Tag für 1 und 3 Tage; sowie drei zusätzliche nicht-hepatotoxische Substanzen: Cefuroxim (Sigma, St. Louis, USA), in einer Konzentration von 250 mg/kg/Tag für 1, 3, 7 und 14 Tage, Nifedipin (Sigma, St. Louis, USA), in einer Konzentration von 3 mg/kg/Tag für 1, 3, 7 und 14 Tage, und Propranolol (Sigma, St. Louis, USA), in einer Konzentration von  
25 40 mg/kg/Tag für 1, 3, 7 und 14 Tage.

Die Dosierungen der Karzinogene wurde so ausgewählt, dass ein Lebertumor lediglich unter der Bedingung einer langfristigen Verabreichung entsteht, so dass eine kurzfristige Verabreichung dieser Karzinogene in einem Bereich von 14 Tagen lediglich einen reversiblen toxischen Stress auf die Ratten ausübt. Für jede Verabreichungsgruppe wurde einer entsprechenden Gruppe von  
30 Kontrolltieren in gleicher Weise Lösemittel appliziert.

Nach den pro Substanz angegebenen Tagen der Applikation wurde jeweils die Gesamt-RNA der Leber von jeweils 3 gleichbehandelten Versuchstieren mittels RNAeasy 96 well-Kits (Qiagen) isoliert. Die Analyse der RNA Expression erfolgte mit der Affymetrix Gene Chip Microarray

Plattform(Affymetrix Inc., Santa Clara, USA) nach einem Standardprotokoll ("GeneChip Sample Cleanup Module, Section 2: Eukaryotic Target Preparation, Affymetrix 701194 Rev.1, 2002). Die einzelnen Schritte sind im folgenden kurz beschrieben. 5 µg der Gesamt-RNA wurden wie vorgegeben mit dem cDNA Doppelstrang Synthese Kit, (Life Technologies, Karlsruhe) in doppelsträngige cDNA umgeschrieben. Von der gereinigten cDNA wurde anschließend in einer in vitro Transkriptionsreaktion mit dem ENZO Bio Array high Yield RNA transcript Labeling Kit, (Affymetrix Inc., Santa Clara, USA) biotinylierte copy-RNA (cRNA) hergestellt. Nach einer Fragmentierung wurden 15µg der biotinylierten cRNA mit RAE230A Microarrays hybridisiert (Affymetrix Inc., Santa Clara, USA).

10 Nach der Hybridisierung für 16 Stunden wurden die Arrays entsprechend den Angaben des Herstellers gewaschen und mit Phycoerythrin-markiertem Streptavidin (Molecular Probes, Eugene, USA), gefärbt. Die Phycoerythrin-Fluoreszenz wurde anschließend in einem Agilent Gene Array Scanner (Agilent, Palo Alto, USA) gelesen.

Der RAE230A Microarray repräsentiert 15.866 sogenannte "Probe Sets". Diese entsprechen 14.280 rattenspezifischen Unigene Clustern, die wiederum größtenteils einzelnen Rattengenomen entsprechen. Die vom Scanner ausgegebenen Rohdatenfiles (DAT) wurden mit Hilfe der Software Microarray Suite 5.0 (MAS5) der Firma Affymetrix durch Hintergrundkorrektur und Mittelung der Fluoreszenzwerte aller 36 pixels pro Oligonukleotid Set in CEL files umgewandelt. Danach erfolgte eine Qualitätskontrolle der Microarrays mit der Software Expressionist Refiner der Firma Genedata AG (Basel, Schweiz). Diese kann Fluoreszenzgradienten und helle oder dunkle Flecken pro Microarray erkennen und korrigieren. In den CEL files wird ein "Probe Set" durch 11 Paare von "Perfect Match (PM)" und "Mismatch (MM)" Oligonukleotid Sets dargestellt, wobei in den MM Oligonukleotiden ein Nukleotid in der Mitte ausgetauscht ist, somit nicht mehr mit der passenden cRNA des durch das PM repräsentierte Gen hybridisieren kann, und damit ein Maß für eine unspezifische Hintergrundhybridisierung darstellt.

Im folgenden wurden dann die Intensitätswerte der einzelnen PMs und MMs pro "Probe Set" mit zwei verschiedenen Algorithmen zu einem Intensitätswert verrechnet. Diese Algorithmen, genannt MAS5 und GCRMA, führen zu etwas unterschiedlichen Intensitätswerten im Niedrig-Expressionsbereich. Die daraus resultierenden zwei Sets von Datenfiles, mit einem Intensitätswert pro "Probe Set" wurden dann wie im folgenden Beispiel beschrieben, eingesetzt.

Insgesamt wurden Microarrays von 138 Lebergewebeproben hybridisiert, wobei die Proben in Gruppen entsprechend Leberproben von Tieren, denen genotoxische Karzinogene (Gruppe 1), nicht-genotoxischen Karzinogene (Gruppe 2), nicht-hepatotoxischen Karzinogene (Gruppe 3)

appliziert wurden, und den jeweiligen Kontrollen der Genexpression vor der Applikation des Karzinogens (Gruppe 0), eingeteilt wurden.

### Beispiel 2

Berechnung der Änderung der Genexpression anhand des linearen Modells

- 5 Für die Erstellung des Modells wurden die mittels Affymetrix gemäß Beispiel 1 bestimmten 4.000 höchst exprimierenden Gene verwendet. Die Auswahl erfolgte, indem die mittlere Expression jedes Gens berechnet wurde und dann die 4000 Gene mit der höchsten mittleren Expression ausgewählt wurden. Diese Auswahl erfolgte, um Fehler in der Auswertung von Expressionsdaten bei niederen Expressionswerten zu vermeiden..

- 10 Für jedes der 4.000 Gene  $i$  wurde die logarithmische Expressionsrate  $x_i$  einzeln berechnet.

Hierzu wurde für alle Daten, die mit Hilfe von GCRMA aus den Rohmessdaten gewonnen werden, der natürliche Logarithmus mit Hilfe von Matlab berechnet.

- 15 Weiterhin wurden die erhaltenen Daten für jedes Gen stratifiziert. Hierzu wurde für jedes Gen der Mittelwert der Expression für jede Substanzgruppe berechnet. Danach wurde für jedes Gen und jeden Expressionswert der jeweilige Mittelwert subtrahiert. Hierdurch wurde erreicht, dass nur noch die Schwankungen um den jeweils durch die Mittelwerte beschriebenen stationären Zustand order Steady state berücksichtigt wurden.

Bei der Bestimmung des jeweiligen stationären Zustands oder Steady State wurde für jedes Gen über jede Substanzgruppe 0, 1, 2 und 3 der Mittelwert berechnet.

- 20 Hierdurch wurde für jedes Gen  $i$  ein Wert  $x_i$ , der die mittlere Verschiebung der Genexpression der  $i$ -ten Komponente als Reaktion auf den toxischen Stress wiedergibt, erhalten. Zusätzlich wurde für jedes Gen  $i$  und jede Gewebeprobe der stratifizierte Expressionswert  $\xi_i$  berechnet, indem von allen Expressionswerten des Gens  $i$  in den Geweben der Stressgruppe der Mittelwert der Expression des Gens  $i$  in dieser Gewebegruppe subtrahiert wurde. Diese Werte geben das Rauschen um den  
25 Mittelwert der jeweiligen Gruppe jeder Substanzgruppe 0, 1, 2 und 3 an. Dieses Rauschen wird zum einen von Messfehlern, zum anderen von der Biodiversität der Reaktion der Gene auf den jeweiligen toxischen Stress und zusätzliche, stochastisch schwankende Umweltbedingungen, erzeugt.

- 30 Aus den Werten  $\xi_i$  wurden je Gen die Standardabweichung  $\sigma_i$  über die verwendeten 138 Proben mit Hilfe von Matlab berechnet.

Aus bekannten Werten von der mittleren Verschiebung  $x_i$  und  $\sigma_i$  wurde der Term  $x_i/\sigma_i$  für die Gene berechnet. Dieser Term gibt effektive Verschiebung der Genexpression der einzelnen Gene durch die Störung an.

Aus den erhaltenen Werten von  $x_i/\sigma_i$  für die 4000 Gene wurden die 100 höchst signifikanten 5 Gene mit den höchsten Werten von  $x_i/\sigma_i$  ausgewählt.

Für diese 100 höchst signifikanten Gene wurden die Gewichte  $w_i$  durch Optimierung mit Hilfe eines genetischen Algorithmus berechnet. Diese Prozedur wird unten beschrieben. Aus diesen Gewichten wurde gemäß

$$\xi_u = \sum_i w_i \xi_i$$

10  $\xi_u$  berechnet und mit dem bekannten  $\xi_i$  wurde dann gemäß Gleichung (IV) der paarweise Korrelationskoeffizient  $\text{cor}(\xi_i, \xi_u)$  berechnet.

In der folgenden Tabelle 1 sind die Werte für  $x_i/\sigma_i$  und  $\text{cor}(\xi_i, \xi_u)$  exemplarisch für die 100 höchst exprimierten Gene angegeben:

Gen-Nummer	$x_i/\sigma_i$ für die 100 höchstexprimierenden Gene	$\text{cor}(\xi_i, \xi_u)$ für die 100 höchstexprimierenden Gene	Gen-Nummer	$x_i/\sigma_i$ für die 100 höchstexprimierenden Gene	$\text{cor}(\xi_i, \xi_u)$ für die 100 höchstexprimierenden Gene
1	-0,8639	-0,2284	51	-1,6086	-0,1596
2	-1,7449	-0,2937	52	-0,8255	-0,2469
3	-0,3352	-0,1256	53	-1,2481	-0,1669
4	-1,714	-0,1832	54	-1,7704	-0,2794
5	-0,1267	0,054	55	-0,8749	-0,1012
6	-1,1887	-0,1871	56	1,1776	0,204
7	-0,5797	-0,0272	57	-1,4196	-0,2213
8	-1,1887	-0,1375	58	-1,5482	-0,1247
9	-0,9122	-0,0954	59	-1,2607	-0,1632
10	-0,7818	-0,1221	60	-1,661	-0,249
11	1,0403	0,1477	61	-3,182	-0,4786
12	-0,621	-0,005	62	-0,5108	-0,1255
13	-1,0258	-0,1489	63	0,3719	0,092
14	-2,0452	-0,2533	64	1,6891	0,2705

Gen- Nummer	$x_i/\sigma_i$ für die 100 höchstexprimie- renden Gene	$\text{cor}(\xi_i, \xi_n)$ für die 100 höchst- exprimierenden Gene	Gen- Nummer	$x_i/\sigma_i$ für die 100 höchstexprimie- renden Gene	$\text{cor}(\xi_i, \xi_n)$ für die 100 höchst- exprimierenden Gene
15	-1,5043	-0,0387	65	-0,7853	-0,1772
16	-1,8747	-0,2316	66	-0,0616	0,0251
17	1,2427	0,0753	67	-1,6085	-0,2457
18	-1,1158	-0,2487	68	-1,1772	-0,228
19	-0,0269	0,0349	69	-1,8573	-0,202
20	-1,5387	-0,2411	70	1,4588	0,2035
21	-0,5044	0	71	-0,1823	0,0684
22	1,4232	0,1759	72	0,2329	0,1671
23	-1,2783	-0,1777	73	1,3752	0,1567
24	-2,0932	-0,3754	74	-1,3919	-0,2328
25	-1,9516	-0,261	75	-2,486	-0,3218
26	-0,8018	-0,1673	76	-1,616	-0,2251
27	-1,5668	-0,2338	77	-1,616	-0,2251
28	-2,731	-0,2212	78	-0,1054	-0,0522
29	-2,8363	-0,3401	79	1,1247	0,1754
30	-1,0813	0,0704	80	-0,8774	-0,1094
31	0,0119	-0,0596	81	-0,0144	0,0008
32	0,964	0,1351	82	-1,709	-0,0839
33	-1,1782	-0,0393	83	-1,8448	-0,2745
34	-1,6021	-0,17	84	-2,8029	-0,2393
35	-0,9161	-0,1772	85	1,712	0,3029
36	-1,6307	-0,3445	86	0,8732	0,1756
37	-0,634	-0,0916	87	-2,7089	-0,251
38	-0,1102	-0,0148	88	-1,7333	-0,2831
39	-0,1269	0,0543	89	-0,9931	-0,0826
40	-1,9546	-0,3756	90	-0,9297	-0,0934
41	-0,3329	0,0894	91	0,8024	0,1281
42	0,1357	-0,1004	92	0,8872	0,066
43	-0,33	0,1339	93	-1,0377	-0,1278
44	-0,5336	-0,012	94	-0,3729	-0,1597
45	-0,0215	-0,0694	95	-0,5099	-0,062

Gen- Nummer	$x_i/\sigma_i$ für die 100 höchstexprimie- renden Gene	cor ( $\xi_i$ , $\xi_u$ ) für die 100 höchst- exprimierenden Gene	Gen- Nummer	$x_i/\sigma_i$ für die 100 höchstexprimie- renden Gene	cor ( $\xi_i$ , $\xi_u$ ) für die 100 höchst- exprimierenden Gene
46	0,5651	0,1144	96	1,3229	0,1239
47	-0,456	-0,0907	97	-2,1548	-0,2142
48	-1,5579	-0,2523	98	-2,1819	-0,3201
49	-1,406	-0,2453	99	-0,3307	-0,032
50	-1,6404	-0,2383	100	2,9326	0,4455

Die Berechnung von  $\xi_u$  erfolgte folgendermassen:

Die Berechnungen wurden mit Hilfe der 4000 höchst exprimierenden Genen durchgeführt, wobei jeweils die 100 signifikantesten Gene als Trainingsdatensatz zur Berechnung der Parameter  
5 verwendet werden, die restlichen 3900 Gene als Testdatensatz zum Test der Modellqualität mit den erhaltenen Parametern.

Um die Stabilität des Modells zu verbessern, wurde nur ein Teil von ca. 30 Genen aus diesen 100 Gene für die Modellierung verwendet. Um diesen Teil optimal zu bestimmen, wurde unter Verwendung des genetischen Algorithmus der Vektor  $\xi_u$  optimiert, indem schrittweise diese  
10 Teilmenge an Genen mit Hilfe des genetischen Algorithmus so ausgewählt wurde, dass das Modell einen minimalen Fehler aufwies.

Die optimale Auswahl dieser Gengruppe erfolgte mit Hilfe eines genetischen Algorithmus wie in der Literatur beschrieben. Es wurden hierzu 20 Gengruppen mit je 20 Genen gebildet. Für jede Gengruppe wurden dann die Gewichte  $w_i$  durch Auflösen der Gleichung (V) nach Einsetzen der  
15 Gleichung (VI) mit Hilfe der linearen Algebraroutinen von Matlab unter Benutzung der 100 höchst signifikanten Gene errechnet. Dann wurde für jede Gengruppe mit den berechneten Gewichten  $w_i$  nach Gleichung (VI) bestimmt und mit Hilfe von Gleichung (V) und der oben genannten Formel für  $\xi_u$  die Vorhersagewerte für die übrigen 4000 Gene berechnet. Das mittlere Fehlerquadrat der Abweichung dieser Vorhersagewerte von den gemessenen Werten ergab das Maß für die Qualität  
20 des Modells, das durch die jeweilige Gengruppe bestimmt ist. Dann wurden wie bei genetischen Algorithmen üblich die 20 Gengruppen durch Rekombination und Mutation variiert und mit den variierten Gengruppen die Berechnung der Modellparameter und der jeweiligen Modellqualität erneut ausgeführt. Diese Prozedur wurde solange wiederholt, bis sich keine Verbesserung mehr erzielen ließ. Nach 200 Wiederholungen wurde keine signifikante Verbesserung der Prognose-  
25 fähigkeit des Modells mehr erzielt.

Dieser optimierte Vektor  $\xi_u$  wurde dann verwendet, um gemäß der Gleichung (IV) die Änderung der Genexpression aller Gene des Netzwerks zu berechnen.

In der folgenden Tabelle 2 sind die Werte für  $\xi_u$ , das als Ergebnis der Optimierung erhalten wurde, für die 138 verwendeten Gewebeproben angegeben:

Gewebe- Nummer	Vektor $\xi_u$	Gewebe- Nummer	Vektor $\xi_u$	Gewebe- Nummer	Vektor $\xi_u$
1	76,5569	47	36,0453	93	-18,4466
2	-14,5742	48	-14,2015	94	-85,8512
3	288,1599	49	145,4589	95	10,727
4	11,1768	50	-160,644	96	-109,0306
5	230,3513	51	77,3361	97	-43,5056
6	191,2853	52	48,4672	98	89,0143
7	188,1156	53	-40,5311	99	-116,9526
8	291,1027	54	-74,2292	100	-102,3417
9	224,6252	55	47,0955	101	-56,6384
10	-53,294	56	-50,7783	102	-167,215
11	-90,4583	57	-107,2944	103	9,239
12	-294,7351	58	-459,4381	104	-42,8732
13	274,2629	59	-581,0783	105	-68,8991
14	-56,5562	60	116,1542	106	-72,9573
15	-28,1301	61	177,4406	107	33,4551
16	167,6595	62	149,8353	108	-30,4143
17	-137,847	63	58,9269	109	-186,0175
18	-77,9698	64	167,4023	110	-13,3843
19	-54,7617	65	-59,1586	111	-25,0929
20	-169,1818	66	-605,2145	112	-150,191
21	-5,0533	67	316,2251	113	-186,7943
22	15,2488	68	322,8739	114	16,8619
23	-82,9799	69	-51,0424	115	79,3224
24	-73,3627	70	245,5574	116	91,981
25	-268,3438	71	66,4274	117	-172,7753
26	-27,8142	72	42,202	118	-44,9154
27	-31,3407	73	-21,7779	119	-46,1011
28	-234,3951	74	91,325	120	136,5539

Gewebe- Nummer	Vektor $\xi_u$	Gewebe- Nummer	Vektor $\xi_u$	Gewebe- Nummer	Vektor $\xi_u$
29	208,0049	75	-52,6885	121	94,5613
30	98,5644	76	-57,2132	122	-121,9597
31	-17,0821	77	-149,6873	123	-211,345
32	3,8032	78	78,5563	124	-95,7291
33	-1,7166	79	448,4771	125	23,3157
34	13,7851	80	-185,6028	126	50,9724
35	36,9275	81	-56,3119	127	198,6063
36	-275,6066	82	113,9029	128	227,2184
37	83,9284	83	183,2596	129	101,7276
38	-7,4295	84	107,4858	130	29,5541
39	-43,6217	85	128,9119	131	-62,2693
40	77,6214	86	146,4095	132	119,2673
41	-36,2371	87	-100,1825	133	-224,152
42	30,5607	88	83,4926	134	153,3749
43	0,5632	89	21,8313	135	341,7285
44	-99,5823	90	-312,5623	136	126,7139
45	-33,3024	91	78,934	137	-107,1419
46	18,2819	92	-75,366	138	28,559

Dieser optimierte Vektor  $\xi_u$  wurde dann verwendet, um gemäß der Gleichung (IV) die Änderung der Genexpression der 4.000 Gene des Netzwerks zu berechnen.

Es zeigte sich, dass die mit dem bereitgestellten linearen Modell bestimmte Änderung der Genexpression aller Gene des Netzwerks eine gute Übereinstimmung mit den gemessenen Daten zeigt. So zeigte eine Auftragung von  $x_i/\sigma_i$  gegen  $\text{cor}(\xi_i, \xi_u)$  dass die durch die reversible Störung regulierten Gene, insbesondere durch die Störung durch nicht-genotoxische Kanzerogene, eine gute Übereinstimmung mit dem linearen Modell zeigten.

Es zeigte sich weiter, dass die Gene, die dicht am biologisch vermuteten Angriffspunkt lagen, tatsächlich einen hohen Korrelationskoeffizienten mit  $\xi_u$  hatten. Darüber hinaus zeigt es sich, dass keine signifikanten systematischen Abweichungen vom Modell auftraten, so dass die im Experiment verursachten Störungen durch nicht-genotoxische Kanzerogene keinen signifikanten nichtlinearen Anteil hatten und somit als reversibel klassifiziert werden konnten.

**Patentansprüche**

1. Verfahren zur Bestimmung des Verhaltens wenigstens eines biologischen Systems nach einer reversiblen Störung umfassend die folgenden Schritte:
  - 5 (a) Bereitstellen wenigstens eines biologischen Systems, wobei das biologische System ein biologisches Netzwerk umfassend eine Vielzahl von biologischen oder biochemischen Komponenten, die eine Aktivität aufweisen, umfasst;
  - (b) Bereitstellen eines linearen Modells zur Beschreibung des Verhaltens des Netzwerks des biologischen Systems;
  - 10 (c) Bestimmen der Aktivität der biologischen oder biochemischen Komponenten des biologisches Netzwerks;
  - (d) reversibles Stören der Aktivität wenigstens einer der biologischen oder biochemischen Komponenten, wobei eine Reaktion des biologisches Netzwerks erzeugt wird, die durch die Änderung der Aktivität wenigstens einer oder mehrerer der biologischen oder biochemischen Komponenten ausgebildet wird;
  - 15 (e) Bestimmen der Aktivität der biologischen oder biochemischen Komponenten des biologisches Netzwerks nach Ausüben der reversiblen Störung, sobald die Komponenten des Netzwerks die Reaktion auf die Störung vollzogen haben;
  - (f) Bestimmen der Änderung der Aktivität wenigstens einer biologischen oder biochemischen Komponente des biologischen Netzwerks als Reaktion auf die  
20 reversible Störung;
  - (g) Berechnen des Verhaltens des biologischen Netzwerks anhand des bereitgestellten linearen Modells zur Beschreibung des Verhaltens des biologischen Netzwerks und der in Schritt (f) bestimmten Änderung der Aktivität der biologischen oder biochemischen Komponente(n) des biologischen Netzwerks nach der reversiblen  
25 Störung, unter Berücksichtigung der Biodiversität der Reaktion der biologischen oder biochemischen Komponente(n); und
  - (h) optional Vergleichen zwischen der gemäß Schritt (f) bestimmten Änderung der Aktivität der einzelnen Komponenten und dem gemäß Schritt (g) berechneten Verhalten des biologischen Netzwerks anhand des bereitgestellten linearen Modells,  
30 wobei man erwartet, dass eine Übereinstimmung des berechneten Verhaltens mit

den in Schritt (f) bestimmten Änderung der Aktivität der biologischen oder biochemischen Komponente(n) besteht.

2. Verfahren nach Anspruch 1, **dadurch gekennzeichnet**, dass

das bereitgestellte lineare Modell umfasst:

- 5 - einen Vektor, der die Bestimmung der Änderung der Aktivität wenigstens einer biologischen oder biochemischen Komponente des biologischen Netzwerks als Reaktion auf die reversible Störung umfasst,
- eine Matrix, die die Parameter enthält, die die Reaktion der Komponenten auf die Störung beschreiben, und
- 10 - einen Vektor, der die Störung beschreibt.

3. Verfahren nach Anspruch 1 oder 2, **dadurch gekennzeichnet**, dass

der Schritt des Berechnens des Verhaltens des biologischen Netzwerks beinhaltet, dass man eine Matrix, die die Parameter enthält, die die Reaktion der Komponenten auf die Störung beschreiben, durch eine  $n \times n$ -Matrix beschreibt, wobei  $n$  der Anzahl der  
15 Komponenten entspricht.

4. Verfahren nach einem der vorherigen Ansprüche, **dadurch gekennzeichnet**, dass

man die Matrix durch eine Projektion der Daten der Änderung der Aktivität auf ihre Eigenvektoren mit Hilfe der Korrelationskoeffizienten von Komponentenpaaren des biologischen Netzwerks beschreibt.

20 5. Verfahren nach einem der vorherigen Ansprüche, **dadurch gekennzeichnet**, dass

der Vektor, der die Störung beschreibt, einen Rauschanteil aufweist, der die Biodiversität der Reaktion der biologischen oder biochemischen Komponente(n) beschreibt.

6. Verfahren nach einem der vorherigen Ansprüche, **dadurch gekennzeichnet**, dass

25 die Biodiversität eine biologische Variationen ist ausgewählt aus der Gruppe umfassend natürliche Variation einer Aktivität einer Komponente oder eines Netzwerks, eine natürliche Variation eines biologischen Systems und/oder eine Variation der biologischen Reaktionen eines Systems auf Umgebungsfaktoren, die es erlaubt, das Modell mit Hilfe der

durch die Biodiversität erzeugten Variationen ohne systematische Experimente zu bestimmen.

7. Verfahren nach einem der vorherigen Ansprüche, **dadurch gekennzeichnet**, dass  
5 die Störung ein externer Stress ist, vorzugsweise ausgewählt aus der Gruppe umfassend toxischer Stresse, bevorzugt ausgewählt aus der Gruppe umfassend Stress durch nicht-  
genotoxische oder genotoxische Hepatokarzinogene, Wärmestress, Hunger, Stress durch  
Applikation eines pharmazeutischen Wirkstoffs, einer Chemikalie und/oder eines Medi-  
kaments.
8. Verfahren nach einem der vorherigen Ansprüche, **dadurch gekennzeichnet**, dass  
10 man das biologische System auswählt aus der Gruppe umfassend Zelle(n), Gewebe,  
Organ(e) und/oder Organismus.
9. Verfahren nach einem der vorherigen Ansprüche, **dadurch gekennzeichnet**, dass  
die biologische Komponente ein Gen ist.
10. Verfahren nach einem der vorherigen Ansprüche, **dadurch gekennzeichnet**, dass  
15 man die biochemische Komponente auswählt aus der Gruppe umfassend RNA, DNA,  
Metabolit und/oder Protein.
11. Verfahren nach einem der vorherigen Ansprüche, **dadurch gekennzeichnet**, dass  
20 die Störung eine direkte Änderung der Aktivität einer Anzahl an Komponenten eines  
Netzwerks im Bereich von  $\geq 1$  Komponente bis alle Komponenten, entsprechend  $\leq 100$  %  
der Komponenten, bezogen auf 100 % Komponenten, bevorzugt im Bereich bis  $\leq 10$  % der  
Komponenten, vorzugsweise im Bereich bis  $\leq 5$  % der Komponenten, bezogen auf 100 %  
Komponenten, hervorruft.
12. Verfahren nach einem der vorherigen Ansprüche, **dadurch gekennzeichnet**, dass  
25 man die Änderung der Aktivität wenigstens einer biologischen oder biochemischen  
Komponente mittels Verfahren ausgewählt aus der Gruppe umfassend semiquantitative  
RT-PCR, Northern Hybridisierung, Differentialdisplay, substraktiver Hybridisierung,  
subtrahierten Bibliotheken, cDNA Arrays und/oder Oligo-Arrays bestimmt.
13. Verfahren nach einem der vorherigen Ansprüche, **dadurch gekennzeichnet**, dass

man in einem weiteren Schritt erhält, dass eine statistisch signifikante Regulation der Aktivität einer oder mehrerer Komponente(n) gemäß der in Schritt (f) bestimmten Änderung der Aktivität und dem gemäß Schritt (g) berechneten Verhalten der Komponente im Netzwerk vorliegt.

- 5 14. Verfahren nach einem der vorherigen Ansprüche, **dadurch gekennzeichnet**, dass
- man die Schritte (a) bis (h) für wenigstens zwei reversible Störungen und optional wenigstens zwei Systeme wiederholt, und in einem weiteren Schritt des Vergleichs erhält, dass eine statistisch signifikante Regulation der Aktivität einer oder mehrerer Komponente(n) gemäß der in Schritt (f) bestimmten Änderung der Aktivität und dem gemäß
- 10 Schritt (g) berechneten Verhalten der Komponente in Bezug auf verschiedene Arten von Störungen vorliegt, die eine Klassifikation der Störung anhand des Auftretens der statistisch signifikanten Regulation der Komponente(n) erlaubt.
15. Verfahren nach einem der vorherigen Ansprüche, **dadurch gekennzeichnet**, dass
- man in Schritt (h) feststellt, dass eine statistisch signifikante Abweichung einer oder
- 15 mehrerer Komponente(n) der gemäß Schritt (f) bestimmten Änderung der Aktivität und dem gemäß Schritt (g) berechneten Verhalten der Komponente(n) im Netzwerk vorliegt, die anzeigt, dass diese Komponente(n) nicht dem bereitgestellten linearen Modell unterliegt.
16. Verfahren nach einem der vorherigen Ansprüche, **dadurch gekennzeichnet**, dass
- 20 man die statistische Signifikanz mittels eines Signifikanztests vorzugsweise ausgewählt aus der Gruppe umfassend T-Test, Z-Test und/oder chi-Quadrat-Test bestimmt.
17. Verfahren nach einem der vorherigen Ansprüche umfassend die folgenden Schritte:
- (a) Bereitstellen eines Organismus, der ein Gewebes enthält, das ein biologisches Netzwerk umfassend eine Vielzahl von Genen umfasst;
- 25 (b) Bereitstellen eines linearen Modells zur Beschreibung der Änderung der Genexpression des Netzwerks;
- (c) Bestimmen der Grundexpression der Gene;

- (d) Ausüben eines toxischen Stresses, bevorzugt Applikation eines pharmazeutischen Wirkstoffs, vorzugsweise eines Karzinogens, wobei eine Änderung der Genexpression erzeugt wird;
- (e) Bestimmen der Genexpression nach Applikation des toxischen Stresses, bevorzugt des pharmazeutischen Wirkstoffs, vorzugsweise des Karzinogens, sobald die Gene des Netzwerks die Reaktion auf den Stress vollzogen haben;
- (f) Bestimmen der Änderung der Expression wenigstens eines Gens nach Ausüben des toxischen Stresses, bevorzugt Applikation des pharmazeutischen Wirkstoffs, vorzugsweise des Karzinogens;
- (g) Berechnen der Änderung der Genexpression aller Gene des Netzwerks anhand des bereitgestellten linearen Modells zur Beschreibung des Verhaltens des biologischen Netzwerks aus der bestimmten Änderung der Expression wenigstens eines Gens unter Berücksichtigung der Biodiversität der Änderung der Genexpression; und
- (h) optional Vergleichen der gemäß Schritt (f) bestimmten Änderung der Expression wenigstens eines Gens und der gemäß Schritt (g) berechneten Änderung der Genexpression der Gene des Netzwerks anhand des bereitgestellten linearen Modells, wobei man erwartet, dass eine Übereinstimmung berechneten Änderung der Genexpression mit den in Schritt (f) bestimmten Änderung der Expression wenigstens eines Gens besteht.
18. Verfahren zur Bestimmung der Änderung der Genexpression in einem Gewebe nach Anspruch 17, **dadurch gekennzeichnet**, dass
- man die Expression einer Anzahl Gene im Bereich von  $\geq 1$  Gen bis  $\leq 5.000$  Gene, bevorzugt im Bereich von  $\geq 2$  Gene bis  $\leq 1000$  Gene, vorzugsweise im Bereich von  $\geq 5$  Gene bis  $\leq 400$  Gene bestimmt.
19. Computerprogrammprodukt mit computerlesbaren Programmmitteln zur Durchführung eines oder mehrerer Schritte des Verfahrens nach einem der vorherigen Ansprüche, wenn das Programm auf einem Computer ausgeführt wird.
20. Computerprogramm zur Ausführung in einem Computersystem mit Softwarekomponenten zur Durchführung eines oder mehrerer Schritte des Verfahrens nach einem der vorherigen Ansprüche, wenn das Programm auf einem Computer ausgeführt wird.

21. Computersystem mit Mitteln zur Ausführung der eines oder mehrerer Schritte des Verfahrens nach einem der vorherigen Ansprüche.

# INTERNATIONAL SEARCH REPORT

International application No  
PCT/EP2007/005712

**A. CLASSIFICATION OF SUBJECT MATTER**  
INV. G06F19/00

According to International Patent Classification (IPC) or to both national classification and IPC

**B. FIELDS SEARCHED**

Minimum documentation searched (classification system followed by classification symbols)  
G06F

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

EPO-Internal, WPI Data, INSPEC

**C. DOCUMENTS CONSIDERED TO BE RELEVANT**

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	<p>GARDNER TIMOTHY S ET AL: "Inferring genetic networks and identifying compound mode of action via expression profiling." SCIENCE (NEW YORK, N.Y.) 4 JUL 2003, vol. 301, no. 5629, 4 July 2003 (2003-07-04), pages 102-105, XP002454759 ISSN: 1095-9203 cited in the application abstract page 102 page 104 figures 1,3</p> <p style="text-align: center;">----- -/--</p>	1-21

Further documents are listed in the continuation of Box C.

See patent family annex.

\* Special categories of cited documents:

- \*A\* document defining the general state of the art which is not considered to be of particular relevance
- \*E\* earlier document but published on or after the international filing date
- \*L\* document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- \*O\* document referring to an oral disclosure, use, exhibition or other means
- \*P\* document published prior to the international filing date but later than the priority date claimed

- \*T\* later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
- \*X\* document of particular relevance: the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
- \*Y\* document of particular relevance: the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.
- \*&\* document member of the same patent family

Date of the actual completion of the international search

15 October 2007

Date of mailing of the international search report

29/10/2007

Name and mailing address of the ISA/

European Patent Office, P.B. 5818 Patentlaan 2  
NL - 2280 HV Rijswijk  
Tel. (+31-70) 340-2040, Tx. 31 651 epo nl,  
Fax: (+31-70) 340-3016

Authorized officer

Hilbig, Matthias

## INTERNATIONAL SEARCH REPORT

International application No  
PCT/EP2007/005712

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	<p>WO 2005/003368 A (SIEMENS AG [DE]; STETTER MARTIN [DE]; DEJORI MATHAEUS [DE]) 13 January 2005 (2005-01-13) abstract page 4, line 26 - page 5, line 32 page 10, line 30 - page 20, line 10 page 28, line 1 - page 29, line 19 claims 1,14 figure 1</p>	1-21
X	<p>DEJORI M ET AL: "ESTIMATION OF ONCOGENES BY BAYESIAN INVERSE MODELING OF GENE-EXPRESSION PATTERNS" INTERNET CITATION, [Online] 2003, XP002320818 Retrieved from the Internet: URL:<a href="http://www.iscb.org/ismb2003/posters/mathaeus.dejori.externalATmchp.siemens.de_109html">http://www.iscb.org/ismb2003/posters/mathaeus.dejori.externalATmchp.siemens.de_109html</a>&gt; [retrieved on 2005-03-10] the whole document</p>	1-21
X	<p>US 5 930 154 A (THALHAMMER-REYERO CRISTINA [US]) 27 July 1999 (1999-07-27) abstract column 1, line 64 - column 2, line 60 column 3, line 31 - column 8, line 4 column 9, line 17 - column 12, line 8 column 13, line 17 - line 61 column 17, line 32 - column 19, line 60 column 21, line 26 - column 22, line 14 column 27, line 21 - column 30, line 60 column 40, line 29 - column 42, line 7 claims 1-4,21-24 figures 1,4-6,12,16-18</p>	1-21

# INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No  
PCT/EP2007/005712

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 2005003368 A	13-01-2005	US 2006177827 A1	10-08-2006
US 5930154 A	27-07-1999	NONE	

**INTERNATIONALER RECHERCHENBERICHT**

Internationales Aktenzeichen

PCT/EP2007/005712

**A. KLASSIFIZIERUNG DES ANMELDUNGSGEGENSTANDES**  
 INV. G06F19/00

Nach der Internationalen Patentklassifikation (IPC) oder nach der nationalen Klassifikation und der IPC

**B. RECHERCHIERTE GEBIETE**

Recherchierter Mindestprüfstoff (Klassifikationssystem und Klassifikationssymbole)  
 G06F

Recherchierte, aber nicht zum Mindestprüfstoff gehörende Veröffentlichungen, soweit diese unter die recherchierten Gebiete fallen

Während der internationalen Recherche konsultierte elektronische Datenbank (Name der Datenbank und evtl. verwendete Suchbegriffe)

EPO-Internal, WPI Data, INSPEC

**C. ALS WESENTLICH ANGESEHENE UNTERLAGEN**

Kategorie*	Bezeichnung der Veröffentlichung, soweit erforderlich unter Angabe der in Betracht kommenden Teile	Betr. Anspruch Nr.
X	GARDNER TIMOTHY S ET AL: "Inferring genetic networks and identifying compound mode of action via expression profiling." SCIENCE (NEW YORK, N.Y.) 4 JUL 2003, Bd. 301, Nr. 5629, 4. Juli 2003 (2003-07-04), Seiten 102-105, XP002454759 ISSN: 1095-9203 in der Anmeldung erwähnt Zusammenfassung Seite 102 Seite 104 Abbildungen 1,3 ----- -/--	1-21

Weitere Veröffentlichungen sind der Fortsetzung von Feld C zu entnehmen  Siehe Anhang Patentfamilie

- \* Besondere Kategorien von angegebenen Veröffentlichungen :
- \*A\* Veröffentlichung, die den allgemeinen Stand der Technik definiert, aber nicht als besonders bedeutsam anzusehen ist
- \*E\* älteres Dokument, das jedoch erst am oder nach dem internationalen Anmeldedatum veröffentlicht worden ist
- \*L\* Veröffentlichung, die geeignet ist, einen Prioritätsanspruch zweifelhaft erscheinen zu lassen, oder durch die das Veröffentlichungsdatum einer anderen im Recherchenbericht genannten Veröffentlichung belegt werden soll oder die aus einem anderen besonderen Grund angegeben ist (wie ausgeführt)
- \*O\* Veröffentlichung, die sich auf eine mündliche Offenbarung, eine Benutzung, eine Ausstellung oder andere Maßnahmen bezieht
- \*P\* Veröffentlichung, die vor dem internationalen Anmeldedatum, aber nach dem beanspruchten Prioritätsdatum veröffentlicht worden ist
- \*T\* Spätere Veröffentlichung, die nach dem internationalen Anmeldedatum oder dem Prioritätsdatum veröffentlicht worden ist und mit der Anmeldung nicht kollidiert, sondern nur zum Verständnis des der Erfindung zugrundeliegenden Prinzips oder der ihr zugrundeliegenden Theorie angegeben ist
- \*X\* Veröffentlichung von besonderer Bedeutung; die beanspruchte Erfindung kann allein aufgrund dieser Veröffentlichung nicht als neu oder auf erfinderischer Tätigkeit beruhend betrachtet werden
- \*Y\* Veröffentlichung von besonderer Bedeutung; die beanspruchte Erfindung kann nicht als auf erfinderischer Tätigkeit beruhend betrachtet werden, wenn die Veröffentlichung mit einer oder mehreren anderen Veröffentlichungen dieser Kategorie in Verbindung gebracht wird und diese Verbindung für einen Fachmann nahelegend ist
- \*Z\* Veröffentlichung, die Mitglied derselben Patentfamilie ist

Datum des Abschlusses der internationalen Recherche	Absenddatum des internationalen Recherchenberichts
15. Oktober 2007	29/10/2007
Name und Postanschrift der Internationalen Recherchenbehörde Europäisches Patentamt, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016	Bevollmächtigter Bediensteter  Hilbig, Matthias

## C. (Fortsetzung) ALS WESENTLICH ANGESEHENE UNTERLAGEN

Kategorie*	Bezeichnung der Veröffentlichung, soweit erforderlich unter Angabe der in Betracht kommenden Teile	Betr. Anspruch Nr.
X	WO 2005/003368 A (SIEMENS AG [DE]; STETTER MARTIN [DE]; DEJORI MATHAEUS [DE]) 13. Januar 2005 (2005-01-13) Zusammenfassung Seite 4, Zeile 26 - Seite 5, Zeile 32 Seite 10, Zeile 30 - Seite 20, Zeile 10 Seite 28, Zeile 1 - Seite 29, Zeile 19 Ansprüche 1,14 Abbildung 1 -----	1-21
X	DEJORI M ET AL: "ESTIMATION OF ONCOGENES BY BAYESIAN INVERSE MODELING OF GENE-EXPRESSION PATTERNS" INTERNET CITATION, [Online] 2003, XP002320818 Gefunden im Internet: URL: <a href="http://www.iscb.org/ismb2003/posters/mathaeus.dejori.externalATmchp.siemens.de_109html">http://www.iscb.org/ismb2003/posters/mathaeus.dejori.externalATmchp.siemens.de_109html</a> > [gefunden am 2005-03-10] das ganze Dokument -----	1-21
X	US 5 930 154 A (THALHAMMER-REYERO CRISTINA [US]) 27. Juli 1999 (1999-07-27) Zusammenfassung Spalte 1, Zeile 64 - Spalte 2, Zeile 60 Spalte 3, Zeile 31 - Spalte 8, Zeile 4 Spalte 9, Zeile 17 - Spalte 12, Zeile 8 Spalte 13, Zeile 17 - Zeile 61 Spalte 17, Zeile 32 - Spalte 19, Zeile 60 Spalte 21, Zeile 26 - Spalte 22, Zeile 14 Spalte 27, Zeile 21 - Spalte 30, Zeile 60 Spalte 40, Zeile 29 - Spalte 42, Zeile 7 Ansprüche 1-4,21-24 Abbildungen 1,4-6,12,16-18 -----	1-21

# INTERNATIONALER RECHERCHENBERICHT

Angaben zu Veröffentlichungen, die zur selben Patentfamilie gehören

Internationales Aktenzeichen

PCT/EP2007/005712

Im Recherchenbericht angeführtes Patentdokument	Datum der Veröffentlichung	Mitglied(er) der Patentfamilie	Datum der Veröffentlichung
WO 2005003368 A	13-01-2005	US 2006177827 A1	10-08-2006
US 5930154 A	27-07-1999	KEINE	