

(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
13. März 2003 (13.03.2003)

PCT

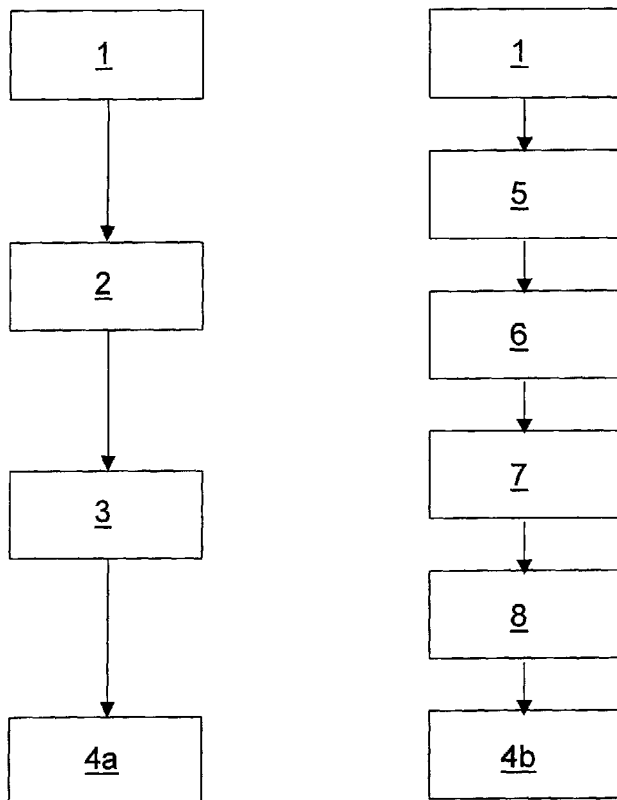
(10) Internationale Veröffentlichungsnummer
WO 03/021478 A2

- (51) Internationale Patentklassifikation⁷: **G06F 17/30** (71) **Anmelder** (für alle Bestimmungsstaaten mit Ausnahme von US): **EUROPROTEOME AG** [DE/DE]; Neuendorfstr. 24a, 16761 Hennigsdorf (DE).
- (21) Internationales Aktenzeichen: PCT/EP02/09735 (72) **Erfinder; und**
- (22) Internationales Anmeldedatum: 30. August 2002 (30.08.2002) (75) **Erfinder/Anmelder** (nur für US): **REYMOND, Marc, Andre** [CH/DE]; Pottkuhlenweg 15, 38350 Helmstedt (DE).
- (25) Einreichungssprache: Deutsch (74) **Anwälte**: **HENGELHAUPT, Jürgen, D.** usw.; Gulde Hengelhaupt Ziebig, Schützenstrasse 15-17, 10117 Berlin (DE).
- (26) Veröffentlichungssprache: Deutsch (81) **Bestimmungsstaaten** (national): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR,
- (30) Angaben zur Priorität: 101 43 712.9 30. August 2001 (30.08.2001) DE

[Fortsetzung auf der nächsten Seite]

(54) **Title**: METHOD AND SYSTEM FOR DATA EVALUATION, CORRESPONDING COMPUTER PROGRAM PRODUCT, AND CORRESPONDING COMPUTER-READABLE STORAGE MEDIUM

(54) **Bezeichnung**: VERFAHREN UND ANORDNUNG ZUR DATENAUSWERTUNG SOWIE EIN ENTSPRECHENDES COMPUTERPROGRAMM-ERZEUGNIS UND EIN ENTSPRECHENDES COMPUTERLESBARES SPEICHERMEDIUM



(57) **Abstract**: The invention relates to a method and a system for data evaluation, to a corresponding computer program product, and to a corresponding computer-readable storage medium, which can be especially used as an internet-based, patient-specific prognosis system. In this case, clinical, pathological and molecular biological data can be integrated, and said data can be combined with relevant prognoses for a particular patient. The system thus enables an oncologist, for example, to decide on an individual treatment on the basis of a specific information pattern. The quality of the prognosis is improved by determining significant and secondary variables, leading to a clear reduction in the quantity of data to be evaluated, to the acceleration of the data evaluation and to the improvement of the prognosis.

(57) **Zusammenfassung**: Die Erfindung betrifft ein Verfahren und eine Anordnung zur Datenauswertung sowie ein entsprechendes Computerprogramm-Erzeugnis und ein entsprechendes computerlesbares Speichermedium, welche insbesondere einsetzbar sind als internetbasiertes patientenspezifisches Prognose-system. Es wird in diesem Falle die Integration von klinischen, pathologischen und molekularbiologischen Daten ermöglicht sowie die Verknüpfung dieser Daten mit relevanten Prognoseaussagen bei einem bestimmten Patienten. Dadurch erlaubt das System beispielsweise einem Onkologen, aufgrund spezifischer

[Fortsetzung auf der nächsten Seite]

WO 03/021478 A2



CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.

SE, SK, TR), OAPI-Patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Veröffentlicht:

— ohne internationalen Recherchenbericht und erneut zu veröffentlichen nach Erhalt des Berichts

(84) Bestimmungsstaaten (regional): ARIPO-Patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), eurasisches Patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), europäisches Patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT,

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 und Anordnung zur Datenauswertung sowie ein
entsprechendes Computerprogramm-Erzeugnis und ein ent-
sprechendes computerlesbares Speichermedium**

Beschreibung

Die Erfindung betrifft ein Verfahren und eine Anordnung zur Datenauswertung sowie ein entsprechendes Computerprogramm-Erzeugnis und ein entsprechendes computerlesbares Speichermedium, welche insbesondere einsetzbar sind als internetbasiertes patientenspezifisches Prognosesystem. Es wird in diesem Falle die Integration von klinischen, pathologischen und molekularbiologischen Daten ermöglicht sowie die Verknüpfung dieser Daten mit relevanten Prognoseaussagen bei einem bestimmten Patienten. Dadurch erlaubt das System beispielsweise einem Onkologen, aufgrund spezifischer Informationsmuster eine individuelle Therapieentscheidung zu treffen.

Die Informationstechnologie (IT) gewinnt in der Medizin zunehmend an Bedeutung. Dennoch wird derzeit die Patientenversorgung als Kernprozeß im Gesundheitswesen nur unzureichend unterstützt. Stattdessen stehen administrative Tätigkeiten im Vordergrund. Das Potential dieser Technologie erlaubt aber die Bereitstellung einer qualitativ besseren Patientenversorgung bei gleich-

zeitiger ökonomischer Nutzung der vorhandenen Ressourcen.

Einen wichtigen Bestandteil einer verbesserten Patientenversorgung bilden verlässliche Prognoseaussagen. Eine
5 Prognose läßt sich jedoch nicht nur anhand allgemeiner Kenntnisse zu Krankheit und Patient erstellen; wichtig sind darüber hinaus auch Informationen zum vorherigen Verlauf der Krankheit bei jedem einzelnen Patienten.
10 Hierbei spielen genaue klinisch-pathologische Daten und Informationen zur Nachsorge eine sehr wichtige Rolle. So muß hier beispielsweise die iterative Natur des Prognosebestimmung berücksichtigt werden.

15 Neben der genauen Kenntnis des Patienten und seiner Krankheit ist es ebenso wichtig, den vorliegenden Fall mit ähnlichen Fällen aus der Vergangenheit an eigenen oder externen Institutionen vergleichen und die damaligen Erfahrungen berücksichtigen zu können. Solche
20 Vergleiche sind jedoch nur mit qualitativ hochwertigen Referenz-Datenbanken realisierbar.

Prognoseaussagen kommt insbesondere bei Krebskrankheiten eine große Bedeutung zu, um die bestmögliche
25 Therapie zu bestimmen. Die Wichtigkeit ergibt sich aus der Tatsache, daß Krebs im Unterschied zu einer Virus-erkrankung, welche in jedem Patienten die gleichen Symptome hervorruft, ein sich entwickelndes und individuelles Krankheitsbild hat. Bei diesen Krankheiten sind
30 für Prognoseaussagen die folgenden Fakten von Bedeutung:

- Patienteninformationen wie Alter, Co-Morbidität, Zuverlässigkeit, usw.,

- Umgebungsinformationen wie Chirurg, initiale Behandlung, Versicherungssystem, Land, usw.,
- Tumorinformationen wie Pathologie, Tumorstaging, Mutationen, Genexpression auf Transkriptom- und Proteom-Ebene.

Um diese „Kunst der Vorhersage“ zu erlernen, ist es wichtig vom Patienten auszugehen. Welche Fragen beschäftigen den Patienten? was will er vom Arzt wissen? Hierbei werden von Patienten - einer Umfrage zufolge - die folgenden Fragen am häufigsten gestellt:

- Inwieweit werde ich mit einer Behandlung geheilt?
- Wie ist die normale Lebensdauer mit einer Behandlung?
- Werde ich sterben, wenn ich nicht behandelt werde?
- Wie schnell wird sich die Krankheit ausbreiten, wenn ich nicht behandelt werde?

Die heute zu Verfügung stehende Staging-Systeme (wie z. B. das TumorNodeMetastasis-System der International Union Against Cancer - UICC) erlauben zwar bereits Aussagen für Patientengruppen, leider aber nicht für einen spezifischen Patienten. Bei der Prognose müssen aber Informationen auf jeden einzelnen Patienten bezogen werden unter Berücksichtigung von dessen spezifischer Situation, wohingegen bei der Diagnose das Spezielle verallgemeinert und übergangen wird. Künftig werden moderne Erkenntnisse aus der Tumor-Genexpressions-Forschung zusätzlich berücksichtigt werden müssen, um diesen Schritt von Diagnose zu patientenspezifischer Prognose - und damit eine individuelle Therapie - zu ermöglichen. Eine Aufgabe, die bisher noch nicht gelöst wurde.

Ein weiteres ungelöstes Problem bei den herkömmlichen Systemen besteht in der Bewältigung der erheblichen Datenmengen, die für eine qualitativ hochwertige Prognose ausgewertet werden müssen. Diese können vom Arzt
5 (Onkologen) allein bei Therapie-Entscheidungen nicht mehr verwaltet werden. Auch die gegenwärtig verfügbare Rechentechnik und die für die Datenauswertung eingesetzten Programme sind nicht geeignet, um diese - insbesondere von den molekularbiologischen Datenbanken
10 bereitgestellten - Datenmengen in vertretbarer Zeit auszuwerten.

Die Nachteile der vorhandenen Lösungen auf diesem Gebiet sind von führenden Krebsverbänden erkannt. Zum
15 Beispiel werden alle Patienten mit kolorektalem Karzinom im UICC-Stadium 3 gemäß Richtlinien der Consensus-Konferenzen nach kurativer Chirurgie adjuvant behandelt, auch wenn nur 40% metachrone Fernmetastasen entwickeln. Dies führt zu überflüssigen Nebenwirkungen
20 für die Patienten, und zu erheblichen Zusatzkosten für das Gesundheitssystem. Auf der anderen Seite werden 8% der Patienten im Stadium I und 14% der Patienten im Stadium II Fernmetastasen entwickeln, und bekommen laut Richtlinien keine adjuvante Chemotherapie, was zu einer
25 erhöhten Krebslethalität führt (Köckerling, Reymond et al, J Clin Oncol, 1998).

Die Etablierung eines internet-basierten medizinischen Informationssystems für Ärzte ist als eine notwendige
30 Unterstützung der ärztlichen Tätigkeit betont worden. Die American Cancer Society (http://www3.cancer.org/cancerinfo/cancer_profiler.asp) und andere Organisationen wie die European School of Oncology (www.cancerworld.org/progetti/cancerworld/start/pagine

/Homeframe.html), die University of Pennsylvania (www.oncolink.upenn.edu/resources/physicians) und andere (z. B. <http://www.cancerhome.com>) bieten schon jetzt Internet-Portale, um z. B. die Onkologen zu unterstützen.

Alle diese Portale geben jedoch nur die aktuellen Richtlinien und Empfehlungen der Experten-Konferenzen wieder und bringen deswegen keine Lösung zur oben
10 erwähnten Problematik. Außerdem können diese Seiten von Patienten aufgerufen werden, was laut dem HON Code of Conduct (HONcode) für medizinische Webseiten im Gesundheitsbereich nicht geschehen darf: „Die Information auf der Website ist so angelegt, daß sie die existierende
15 Arzt-Patienten-Beziehung unterstützt und keinesfalls ersetzt.“ (<http://www.hon.ch/HONcode/German/>).

Weiterhin gibt es bisher kein e-Health-Produkt auf dem Markt, das molekularbiologische Daten mit konventionel-
20 len klinischen und pathologischen Informationen verbindet, um den Arzt bei Therapieentscheidungen für Krebskrankheiten zu unterstützen. Man braucht nur die Vielfalt der möglichen Therapien zu erwähnen, und erhält die Menge an Informationen, die zu bewältigen
25 sind, um die Notwendigkeit eines solchen Systems zu unterstreichen.

Ein Informations-System in der marktvorbereitenden Erprobungsphase, in dem für die Erzeugung der prognostischen Aussage neuronale Netze bzw. regelbasierte
30 Systeme eingesetzt werden, ist bekannt. Hierbei werden einige tausend Datensätze (Sets) von prospektiven Patientendaten für die Auswertung berücksichtigt. Mit einem großen Teil dieser Datensets, die jeweils aus

weit über einhundert Parametern (Variablen) für tausende von Patienten bestehen, wurde das Informationssystem trainiert. Eine zusätzliche Anzahl Datensets von (knapp eintausend) zufällig gewählten Patienten, die
5 nicht zum Trainieren des neuronalen Netzes eingesetzt wurden, wurde zur Prüfung des Systems benützt.

Dieses System ermöglicht eine Voraussage der Überlebenschancen eines Patienten nach kurativer kolorektaler Chirurgie mit einem prädiktiven Wert von etwa 90%. Für
10 die Metastasierung kann das System diese gute prognostische Aussage noch nicht reproduzieren. Hier sind offensichtlich molekularbiologische Zusatzdaten erforderlich, um die Leistung zu verbessern. Am signifikantesten schneiden folgende Variablen ab: Tiefe der
15 Tumordinfiltration, T-Kategorie, tumorfreie Resektionsränder, Grading, venöse und lymphatische Invasion.

Die Aufgabe, die durch die Erfindung gelöst werden soll, besteht darin, ein verbessertes Verfahren zur
20 Datenauswertung bereitzustellen. Ziel ist die Erweiterung der Aussagefähigkeit (die bisher nur für Patientengruppen möglich war) auf Prognosen für einen individuellen Patienten (z. B. hinsichtlich Metastasierungsrisiko, Therapieantwort auf eine Reihe von
25 Chemotherapeutika, und Voraussage von Nebeneffekten). Des weiteren soll durch die Erfindung die Signifikanz der in der Auswertung berücksichtigten Parameter bestimmt und dadurch eine Reduzierung der für die Prognose notwendigen Datenmenge erreicht werden, ohne die
30 Qualität der Prognose zu schmälern.

Diese Aufgabe wird erfindungsgemäß gelöst durch die Merkmale im kennzeichnenden Teil der Ansprüche 1, 13 und 15 bis 25 im Zusammenwirken mit den Merkmalen im

Oberbegriff. Zweckmäßige Ausgestaltungen der Erfindung sind in den Unteransprüchen enthalten.

Ein besonderer Vorteil der Erfindung liegt darin, daß
5 bei dem Verfahren zur Datenauswertung unter Verwendung
von mit Datenbanken gekoppelten Datenverarbeitungseinrichtungen die Datenmengen, die für eine hochwertige
Auswertung, wie z. B. eine medizinische Prognose, be-
rücksichtigt werden müssen, ganz erheblich reduziert
10 werden, wenn der Datenverarbeitungseinrichtung zugeführte
Anfragedaten ausgewertet werden, indem nach vorgebbaren
Regeln und/oder mit Verfahren der künstlichen Intelligenz mit den
Anfragedaten korrespondierende, in der (den) Datenbank(en)
gespeicherte Daten ermittelt
15 werden, die Güte dieser korrespondierenden Daten automatisch
bewertet, anhand dieser Bewertung, der zugehörigen Anfrage-
und/oder korrespondierenden Daten automatisch die Signifikanz
dieser Anfrage- und/oder korrespondierenden Daten für die
Anfrage ermittelt wird,
20 die Ergebnisse der Auswertung, die Gütebewertung und/oder
die Signifikanz der Daten ausgegeben und/oder abrufbar
bereitgestellt werden.

Eine Anordnung zur Datenauswertung ist vorteilhafterweise
25 so eingerichtet, daß sie mindestens einen Prozessor umfaßt,
der (die) derart eingerichtet ist (sind), daß ein Verfahren zur
Datenauswertung durchführbar ist, wobei der Datenverarbeitungseinrichtung
zugeführte Anfragedaten ausgewertet werden, indem nach vorgebbaren
30 Regeln und/oder mit Verfahren der künstlichen Intelligenz mit
den Anfragedaten korrespondierende, in der (den) Datenbank(en)
gespeicherte Daten ermittelt werden, die Güte dieser korrespondierenden
Daten automatisch bewertet, anhand dieser Bewertung, der zugehörigen

gen Anfrage- und/oder korrespondierenden Daten automatisch die Signifikanz dieser Anfrage- und/oder korrespondierenden Daten für die Anfrage ermittelt wird, die Ergebnisse der Auswertung, die Gütebewertung und/oder die Signifikanz der Daten ausgegeben und/oder abrufbar bereitgestellt werden.

Ein Computerprogrammprodukt zur Datenauswertung umfaßt ein computerlesbares Speichermedium, auf dem ein Programm gespeichert ist, das es einem Computer ermöglicht, nachdem es in den Speicher des Computers geladen worden ist, ein Verfahren zur Datenauswertung durchzuführen, wobei die Datenauswertung die Verfahrensschritte gemäß einem der Ansprüche 1 bis 12 umfaßt.

Um eine automatische Datenauswertung durchzuführen, wird vorteilhafterweise ein computerlesbares Speichermedium eingesetzt, auf dem ein Programm gespeichert ist, das es einem Computer ermöglicht, nachdem es in den Speicher des Computers geladen worden ist, ein Verfahren zur Datenauswertung durchzuführen, wobei die Datenauswertung die Verfahrensschritte gemäß einem der Ansprüche 1 bis 12 umfaßt.

Eine Methode zur Nutzung eines Systems zur Datenauswertung besteht darin, daß der Zugang zu dem System durch eine gebührenpflichtige PIN ermöglicht wird, wobei die PIN mit Mitteln zur Erfassung von in das System einzugebenden Daten und/oder mit Mitteln zur Erfassung von Material, welches zur Ermittlung der einzugebenden Daten dient, verbunden ist, und der Nutzer die PIN durch Zahlung einer Gebühr erwirbt.

In einer bevorzugten Ausführungsform der Erfindung ist vorgesehen, daß die Datenauswertung medizinische Prognosen generiert und die Anfragedaten als klinisch-pathologische Daten durch einen Arzt und/oder als biomolekulare Daten durch ein Analyselabor der Datenverarbeitungseinrichtung zugeführt werden. Dabei ist es vorteilhaft, wenn die Anfragedaten und/oder die in der (den) Datenbank(en) gespeicherten Daten krankheitsbezogene Faktoren und/oder patientenspezifische Faktoren und/oder umweltspezifische Faktoren berücksichtigen. Insbesondere werden unter den krankheitsbezogenen Faktoren auch tumorspezifische Faktoren berücksichtigt.

Darüber hinaus hat es sich als vorteilhaft herausgestellt, daß eine Aktualisierung der in der (den) Datenbank(en) gespeicherten Daten Erfahrungsdaten erfolgt, indem Daten über die Therapie und den Krankheitsverlauf zu von der Datenverarbeitungseinrichtung prognostizierten Fällen der Datenverarbeitungseinrichtung zugeführt werden. Ebenso ist vorgesehen, daß eine Aktualisierung der Bewertungsvorschriften in einem iterativen Lernprozeß erfolgt, in welchem Anfrage, Anfragedaten, bei der Auswertung der Anfrage benutzte, in der (den) Datenbank(en) gespeicherten Daten, Ergebnisse der Auswertung und tatsächlich eingetretene Ereignisse berücksichtigt werden.

Von Vorteil ist es insbesondere, wenn in Abhängigkeit von der Bewertung die Menge der in der (den) Datenbank(en) gespeicherten Daten reduziert wird und/oder in Abhängigkeit von der Signifikanz die Zahl der zuzuführenden Anfragedaten und/oder die Menge der in der (den) Datenbank(en) gespeicherten Daten reduziert wird.

In einer bevorzugten Ausführungsform der Erfindung ist außerdem vorgesehen, daß die Auswertung durch Cluster-Analyse, Ähnlichkeiten-Suche, Tendenz-Analyse, Korrespondenzanalyse, aufsteigend hierarchische Klassifikation, Hauptanalyse und/oder Wavelet-Analyse in Verbindung mit aus Irrtums-Modellen generierten Probabilitätswerten und/oder Mitteln der künstlichen Intelligenz wie neuronalen Netzen und/oder regelbasierten Systemen erfolgt.

Darüber hinaus hat es sich als vorteilhaft erwiesen, daß die Datenauswertung die Ermittlung der Eintrittswahrscheinlichkeit von durch die Anfrage und/oder durch die korrespondierenden Daten bezeichneten Ereignissen umfaßt.

Eine Reduktion der zu verarbeitenden Daten kann erreicht werden, indem die Signifikanz von Daten derart ermittelt wird, daß eine Anfrage ein erstes Mal ohne Berücksichtigung dieser Daten und ein zweites Mal unter Berücksichtigung dieser Daten ausgewertet wird, die Bewertungen der Ergebnisse dieser beiden Auswertungen miteinander verglichen und das Maß des Einflusses der Daten relativ zu einer Verbesserung bzw. Verschlechterung der zweiten Bewertung gegenüber ersten Bewertung bestimmt wird und bei einer Verbesserung der zweiten Bewertung gegenüber der ersten Bewertung die Daten als signifikant angesehen und bei künftigen Auswertungen berücksichtigt werden, bei einer Verschlechterung der zweiten Bewertung gegenüber der ersten Bewertung die Daten als nicht signifikant angesehen und für künftige Auswertungen nicht mehr genutzt werden.

Durch die Erfindung wird beispielsweise die Verfügbarkeit klinisch relevanten Wissens zu einem spezifischen medizinischen Problem am Arbeitsplatz des Arztes (Onkologen) erheblich verbessert, indem die
5 Zuführung der Daten, die Anfrage und/oder die Ausgabe der Ergebnisse über das Internet erfolgt. Über das (im Internet erreichbare) Computersystem erhält er Zugriff auf Daten, die aus verschiedenen Datenbanken stammen, aber z. B. durch den Einsatz des Internetstandards XML
10 (eXtensible Markup Language) auf eine einheitliche Basis gestellt wurden. Für ihn sind die unterschiedlichen Formate und Datenstrukturen nicht mehr zu erkennen. Dadurch wird es möglich, die Akzeptanz und den tatsächlichen Einfluß von Genexpressionsdaten auf die
15 Patientenversorgung nachhaltig zu beeinflussen. Durch diese Vereinheitlichung der Sprache bzw. Formate der Daten werden die organisatorischen Voraussetzungen für einen erfolgreichen Einsatz der Erfindung geschaffen.

20 Für den effizienten Einsatz der Erfindung ist eine Anordnung vorgesehen, welche folgendes umfasst:
mindestens eine mit mindestens einer Datenbank gekoppelte Datenverarbeitungseinrichtung,
Mittel zur Datenein- und/oder -ausgabe,
25 nach vorgebbaren Regeln und/oder mit Verfahren der künstlichen Intelligenz arbeitende Mittel zur Ermittlung von in der (den) Datenbank(en) gespeicherten, mit einer Datenverarbeitungseinrichtung zugeführten Anfragedaten korrespondierenden Daten,
30 Mittel zur automatischen Bewertung der Güte der korrespondierenden Daten,
Mittel zur automatischen Ermittlung der Signifikanz der Anfrage- und/oder korrespondierenden Daten für eine Anfrage.

Die Erfindung wird die Schätzung des Metastasierungsrisiko eines individuellen Patienten erlauben, so daß die Indikation einer adjuvanten Chemotherapie gezielter gestellt werden kann. Außerdem wird eine Schätzung der Probabilität der Therapieantwort bei einer Reihe von Chemotherapeutika ermöglicht, so daß mit einer entsprechenden Tumorprofilierung auch Tumorresistenzmuster erkannt werden können. Durch die Selektion der molekularen Targets (auf DNA, RNA und Proteine-Ebenen), die mit einem bestimmten klinischen Outcome (z. B. Metastasierung) verbunden sind, wird eine ganz erhebliche Reduzierung der Daten erreicht, die für eine hochwertige Prognose ausgewertet werden müssen. Es werden nur die Outcome-relevanten Moleküle berücksichtigt, was einen entscheidenden Schritt in Richtung Validierung von potentiellen Drug-Targets im humanen Patienten darstellt. Als weiterer Vorteil der Erfindung ist anzusehen, daß das Risiko von erfolglosen teuren Therapieversuche deutlich gesenkt, die Entwicklungskosten eines Medikamenten reduziert, und damit die Gesundheitskosten verringert werden, da man in der Lage ist, Patientenkollektive zu bestimmen, die für klinische Studien mit einem bestimmten Chemotherapeutikum am besten geeignet sind. Mit Hilfe der Erfindung wird es möglich sein, zu beurteilen, ob eine zusätzliche Therapie zu einer deutlichen Verbesserung der Prognose führt, z. B. verglichen mit dem Pool der Patienten, die mit adjuvanten Therapie behandelt worden sind. Aufgrund dieser spezifischen Informationsmuster wird der Arzt in die Lage versetzt, eine individuelle - d. h. eine auf ein spezifisches Krankheitsbild bzw. Krankheitsstadium bezogene - Therapieentscheidung zu treffen.

So könnten möglicherweise durch den Einsatz der Erfindung prospektiv randomisierte Studien durch evidenzbasierte Daten mit einem hohen Wert ersetzt werden. Dies wäre ein weiterer Vorteil, denn die Durchführung von zahlreichen randomisierten Studien bei einer zunehmender Zahl von Krebstherapien ist mit erheblichen Kosten und organisatorischen Schwierigkeiten verbunden, die somit eingespart werden könnten.

- 10 Die Verwendung eines Verfahrens nach einem der Ansprüche 1 bis 12 erweist sich als vorteilhaft zur
- Auswertung von klinischen, pathologischen und/oder molekulargenetischer Daten,
 - Ermittlung der prognostischen Signifikanz von klinischen, pathologischen und/oder molekulargenetischer
 - 15 Daten,
 - Selektion molekularer Targets,
 - Schätzung des individuellen Risikos, wie beispielsweise des Metastasierungsrisikos, einzelner Patienten,
 - 20 - Schätzung der Probabilität der Therapieantwort z. B. auf Chemotherapeutika und/oder
 - automatischen Generation von Prognose- und/oder Therapievorschlügen.

25

Eine Verbesserung von patientenspezifischen Prognosen kann erwartet werden, indem Daten, Gene, molekulare und/oder genetische Targets genutzt werden, welche durch ein Verfahren nach einem der Ansprüche 1 bis 12, durch eine Anordnung nach einem der Ansprüche 13 oder

30 14, durch ein Computerprogrammprodukt nach Anspruch 15, durch ein computerlesbares Speichermedium nach Anspruch 16 oder eine Verwendung nach Anspruch 17 verfügbar gemacht werden.

Bevorzugt werden deshalb Herstellungsverfahren für diagnostische Anordnungen, welches die Schritte eines Verfahrens nach einem der Ansprüche 1 bis 12 und einen
5 zusätzlichen Schritt umfaßt, in welchem ein diagnostisch effektives Analysewerkzeug, wie z. B. ein RNA- oder Protein-Chip, und/oder eine Kombination von Genen, welche durch ein Verfahren nach einem der Ansprüche 1 bis 12, durch eine Anordnung nach einem der Ansprüche
10 13 oder 14, durch ein Computerprogrammprodukt nach Anspruch 15, durch ein computerlesbares Speichermedium nach Anspruch 16 oder eine Verwendung nach Anspruch 17 verfügbar gemacht wurden, zusammengestellt wird.

15 Desgleichen stellt es einen Vorteil dar, Gene oder Kombination von Genen zu verwenden, welche durch ein Verfahren nach einem der Ansprüche 1 bis 12, durch eine Anordnung nach einem der Ansprüche 13 oder 14, durch ein Computerprogrammprodukt nach Anspruch 15, durch ein
20 computerlesbares Speichermedium nach Anspruch 16 oder eine Verwendung nach Anspruch 17 verfügbar gemacht wurden, für die Vorbereitung einer diagnostischen Zusammenstellung zur Klassifizierung genetisch bedingter Erkrankungen, Tumoren u.ä. und/oder zur Vorhersage
25 genetisch bedingter Erkrankungen und/oder zur Assoziierung von molekular-genetischen Parametern mit klinischen Parametern und/oder zur Identifizierung von Tumoren durch Genexpressionsprofile.

30 Für die Durchführung von Labortests beispielsweise erweist es sich als vorteilhaft, Trägerelemente zu verwenden, auf dem Daten, Gene, molekulare und/oder genetische Targets bereitgestellt werden, welche durch ein Verfahren nach einem der Ansprüche 1 bis 12, durch eine

Anordnung nach einem der Ansprüche 13 oder 14, durch ein Computerprogrammprodukt nach Anspruch 15, durch ein computerlesbares Speichermedium nach Anspruch 16 oder eine Verwendung nach Anspruch 17 verfügbar gemacht werden.

- In einer bevorzugten Ausführungsform des Trägerelements ist vorgesehen, daß das Trägerelement als Chip ausgebildet ist und
- 10 - Daten zum individuellen Risiko, wie z. B. Metastasierungspotential, und/oder
 - Daten zur Therapieantwort z. B. auf Chemotherapeutika und/oder
 - Daten zum Patientenmetabolismus und/oder
 - 15 - Information zur Autoimmunität z. B. gegen einen Tumor bereitstellt.

Vorzugsweise ist das Trägerelement als reproduzierbarer Chip ausgebildet.

Erkenntnisse aus (Internet-)Management und Qualitätssicherung für e-Health-Systeme zeigen, daß die Zielsetzung, die durch die Erfindung umgesetzt wird, eine problemorientierte Darstellung erfordert, die sich nicht allein an einer streng wissenschaftlichen Gliederung orientiert, wie die Genexpressionsdatenvermittlung. Vielmehr ist es notwendig - und die Erfindung erfüllt diese Forderung -, mit Hilfe eines gut strukturierten Aufbaus, der Darstellung von größeren Zusammenhängen und dem Aufzeigen von Methoden und Techniken eine Grundlage zu schaffen, die es dem Arzt (Onkologen) ermöglicht, neben einer individuellen Therapieentschei-

dition außerdem eigene praktische Lösungsansätze auf Basis des vermittelten Wissens bilden zu können.

5 Für die kommerzielle Nutzung der Erfindung erweist es sich als vorteilhaft, wenn der Nutzer die PIN zusammen mit dem (den) Mittel(n) zur Erfassung der Daten und/oder des Materials beim Kauf dieses (dieser) Mittel(s) erwirbt.

10 Eine weitere Möglichkeit der Nutzung des Systems zur Datenauswertung besteht darin, daß ein Vertreiber des Systems zur Datenauswertung mit mindestens einem Kunden Nutzungsvereinbarungen schließt und der (die) Kunde(n) das System für weitere Teilnehmer durch Vergabe gebührenbehafteter PINs nutzbar macht (machen).

Insbesondere ist vorgesehen, daß Referenzlabors, Pharmaunternehmen und/oder Content Provider das System zur Datenauswertung als Kunde nutzen.

20

In einer bevorzugten Variante der kommerziellen Nutzung ist vorgesehen, daß die Gebühren für die Nutzung des Systems zur Datenauswertung über den Kunden erhoben werden und

25 - pro Nutzungsfall und/oder
- prozentual zum Umsatz, den der Kunde mit dem System macht und/oder
- pro vergebener PIN
festgelegt werden.

30

Eine weitere Form der kommerziellen Nutzung besteht darin, daß ein Nutzer des Systems zur Datenauswertung die PIN vom Vertreiber des Systems oder von einem Kunden des Betreibers erwirbt.

Als vorteilhaft erweist es sich, wenn ein Mittel zur Erfassung von Material ein Träger-Chip für die bei einem - erforderlichenfalls reproduzierbaren - Labortest
5 benötigten Proben, wie z. B. ein DNA-Mikroarray, ist.

Hinsichtlich des Datenschutzes erweist es sich als vorteilhaft, wenn eine PIN mit vorgebbaren, in dem System gespeicherten Daten verknüpft ist und die PIN nur
10 Zugang zu diesen mit ihr verknüpften Daten ermöglicht.

Ein weiterer Vorteil besteht in der Verwendung eines Systems zur Datenauswertung nach einem der Ansprüche 1 bis 32 zur Durchführung von Profit- oder Non-Profit-
15 Aktionen durch Ärzte, Patienten und/oder das System betreibende Firmen, wobei eine Aktion durch Teilnehmer und/oder Anbieter des Systems gestartet wird. Solche Aktionen könne etwa den Austausch von Informationen und/oder das Einbringen von Kunden- und/oder Patienten-
20 gruppen beinhalten. Insbesondere ist eine solche Verwendung des Systems zur Datenauswertung sinnvoll, wenn die Aktionen die Schaffung, Unterhaltung und/oder Vermarktung eines Exzellenz-Netzwerkes und/oder die Distribution von Therapien und/oder die Selektion von
25 Patientengruppen für klinische Studien umfassen. Damit kann beispielsweise bei Verwendung des Systems im Internet erreicht werden, daß die Besuchertreue für die entsprechenden Web-Pages gefestigt und/oder ein bestimmter Kundenstamm an das System gebunden wird.

30

Die Erfindung soll nachstehend anhand von zumindest teilweise in den Figuren dargestellten Ausführungsbeispielen näher erläutert werden.

Es zeigen:

- Fig. 1a Schematische Darstellung der Verfahrensschritte bei herkömmlicher Datenauswertung,
- Fig. 1b Schematische Darstellung der Verfahrensschritte bei Datenauswertung gemäß der Erfindung,
- Fig. 2a-d Darstellung des modularen Aufbaus eines medizinischen Informationssystems,
- Fig. 3a-f detaillierte Darstellung des modularen Aufbaus eines medizinischen Informationssystems,
- Fig. 4 Darstellung der beobachteten Überlebenszeiten verschiedener Patientengruppen und Abschätzung nach Kaplan-Meier der Anzahl der Patienten mit fünf Jahren Überlebenszeit,
- 4a Gruppe aller Patienten im UICC-Stadium-III,
- 4b Patienten der Gruppe aus 4a, welche ein zusätzliches Merkmal aufweisen (Gruppe 1), dargestellt im Vergleich zu den Patienten ohne dieses Merkmal (Gruppe 0),
- Fig. 5 Klassifizierung von Patienten innerhalb der drei verschiedenen UICC-Stadien in jeweils zwei Untergruppen von Patienten mit hohem Risiko (high-risk), bzw. geringem Risiko (low risk),
- 5a UICC-Stadium I,
- 5b UICC-Stadium II,
- 5c UICC-Stadium III,
- Fig. 6 ROC-Kurve (ROC = Receiver Operating Characteristic) für eine Voraussage unter Berücksichtigung herkömmlicher Informationen bzw. mit Einbeziehung von zusätzlichen, durch die Datenauswertung nach dem erfindungsgemäßen Verfahren gewonnenen Informationen.

Am Beispiel eines medizinischen Informationssystems für Onkologen, in welchem ein internetbasiertes patientenspezifisches Prognosesystem realisiert wurde, soll Einsatz und Wirkungsweise der Erfindung beschrieben werden.

Das beispielhafte System ist ein internet-basiertes medizinisches Informationssystem, das sich aus Datenbanken, einem Datenreduktionsprogramm und Modulen der künstlichen Intelligenz (neuronales Netzwerk oder regelbasiertes System) zusammensetzt. Es erlaubt die Integration von klinischen, pathologischen und biologischen Daten, und deren Verknüpfung mit relevanten Prognoseaussagen für einen bestimmten Patienten. Damit erlaubt dieses Informationssystem dem Onkologen, aufgrund spezifischer Informationsmuster eine individuelle Therapieentscheidung zu treffen. Die Therapieentscheidungen werden mit Probabilitätsrechnungen unterstützt. Als Prototyp wurde das kolorektale Karzinom gewählt. Das beispielhafte medizinische Informationssystem integriert Daten aus der Transkriptom- und Proteom-Forschung.

Die Anwendung des beispielhaften Informationssystems wird im folgenden beschrieben. Ein Patient wird beim Arzt vorstellig und erkundigt sich nach Therapiemöglichkeiten für seine Krebserkrankung. Nach der Operation schickt der Onkologe die Proben gegen Rechnung an ein Referenzlabor, wo mit beliebigen Laborverfahren wie beispielsweise einem Chip ein oder mehrere Labortests vorgenommen werden, um die notwendigen Genexpressionsanalysen durchzuführen. Mit Einsendung der Proben erhält der Onkologe eine PIN-Nummer, dank der er alle relevanten Daten des Patienten (patientenspezifisch,

umweltspezifisch etc.) in anonymisierter Form in die Datenbank des erfindungsgemäßen Computersystems eintragen kann. Unterdessen trägt das Referenzlabor ebenfalls alle mit der PIN versehenen Tumordaten in diese Datenbank ein. Zu Beginn der Nutzung umfaßt ein Datensatz alle Prognosefaktoren als Variablen, die von der Medizin für die Prognose bei kolorektale Karzinom akzeptiert werden. Dank der PIN-Nummer kann die Korrespondenz der molekularen und klinischen Informationen erkannt werden. Diese kombinierten Informationen werden mit der Datenbank verglichen, und die Patienten mit dem nächstliegenden Informationsmuster und deren Krankheitsverlauf wird herausgesucht. Es wird dabei ein rückwirkendes Fehlerminimierungsverfahren angewendet. Der Arzt (Chirurg oder Onkologe) kann anschließend mit der PIN beim Informationssystem verschiedene Voraussetzungen anfragen. Innerhalb von Minuten erhält dieser dann Angaben zu Metastasierungswahrscheinlichkeit, Resistenzprofile zu verschiedenen Chemotherapeutika, möglicherweise zu Immuntherapeutika oder ähnlichem. Die so gewonnenen Erkenntnisse bilden damit eine wichtige Entscheidungshilfe bei der Therapieentscheidung. Es ist heute nicht vorgesehen, daß Patienten direkten Zugriff auf die Webseite erhalten, die den Zugang zu dem erfindungsgemäßen Computersystem gestattet, jedoch wird diese Möglichkeit für die Zukunft offen gehalten. Später wird der Arzt von der Datenbank regelmäßig über Therapie und Verlauf seines Patienten befragt, diese Daten werden zum iterativen Lernprozeß des Systems gebraucht, so daß sich dieses laufend dem medizinischen Fortschritt anpassen kann. Dadurch werden signifikante und unwesentliche Variablen ermittelt, was zu einer Optimierung der auszuwertenden Datenmenge und zur Verbesserung der Prognose führt. Beispielhaft erfolgt die-

se Unterscheidung zwischen signifikanten und unwesentlichen Variablen dadurch, daß das Informations-System die Genauigkeit der Prognose unter Berücksichtigung der neuen Variablen prüft. Wenn sich diese Genauigkeit verbessert, wird die neue Variable als signifikant berücksichtigt. In den anderen Fällen wird sie als unwesentlich eingestuft und verworfen.

Zum Verständnis der Herkunft und Verknüpfung von Daten muß verdeutlicht werden, daß die heute verfügbaren signifikantesten Daten aus der Klinik, der Pathologie, und der gegebenenfalls schon erfolgten Behandlung es erlauben, eine prognostische Aussage zu treffen. Diese Prognostik wurde im beispielhaften Informationssystem für das kolorektale Karzinom durch moderne Bio-Informatik optimiert. Dieses System hat eine prognostische Leistung erreicht, die bei Hunderten von Patienten durch Kreuzvalidierung genau bestimmt werden konnte. Die Einbeziehung von neuen (molekularbiologischen) Daten erlaubt es, das System erneut zu „trainieren“. Wenn die prognostische Leistung steigt, werden die neuen Daten als prognostisch signifikant bewertet und für weitere Analysen erfordert. Falls das System mit diesem Set neuer Daten nicht besser wird, werden diese Daten ausgeschieden. So können biologische Daten äußerst effizient ausgewählt werden. Mit Data-Mining-Systemen z. B. stehen Auswahl-Verfahren zur Verfügung, welche dies ermöglichen.

In Figur 1a bzw. 1b wird der erfindungsgemäßen Datenauswertung das herkömmliche Verfahren gegenübergestellt. Während bei einem herkömmlichen Verfahren (vgl. Fig. 1a) die Input-Variablen 1 sofort in einem Modul 2 zur Berechnung der Korrelation und anschließend in einem Modul 3 zur multivarianten statistischen Analyse

(bspw. Regressions-Analyse) verarbeitet werden, wird bei dem erfindungsgemäßen Verfahren (vgl. Fig. 1b) nach dem Einlesen der Input-Variablen 1 ein Transformations-schritt 5 durchgeführt. Der Transformationsschritt 5 ist ein wichtiger Schritt des erfindungsgemäßen Verfahrens und dient dazu Nicht-Linearität des Verfahrens zu vermeiden, um den Rechenaufwand klein zu halten. Hierin werden die symbolischen Variablen in geeigneter Form konvertiert. In der anschließenden Feature-Auswahl (Feature-Selection) 6 werden nacheinander die Variablen mit dem höchsten Informationsgehalt ermittelt. Dies wird solange durchgeführt, bis jeder Variable die entsprechende Gewichtung zugeordnet wurde. Als nächster Schritt folgt das Training und die Auswahl des Modells 7. Dieser Verfahrensschritt beinhaltet das Trainieren diverser Modelle mit verschiedenen Input-Variablen und einer Anzahl verborgener Neuronen, die etwa gemäß dem Bayesschen Evidenz-Ansatz berechnet wurden. Das so ermittelte beste Modell 8 kann nun die Prognosen für neue Patienten berechnen, das heißt für neue Eingabedaten einen Ausgabewert bestimmen. Die Modellarchitektur 8 kann immer weiter verbessert und angepaßt werden (das Modell "lernt"). Die Resultate 4b, die bei Verwendung des erfindungsgemäßen Verfahrens erzielt werden, zeichnen sich gegenüber den durch das herkömmliche Verfahren erreichbaren Resultaten 4a durch eine höhere Prognosequalität aus, was vor allem durch die Erstellung patientenspezifischer, individueller Risiko-Profile erreicht wird.

30

Es stellt somit einen erheblichen Vorteil des erfindungsgemäßen Verfahrens zur Datenauswertung besteht dar, daß patientenspezifische, individuelle Risiko-Profile erstellt werden können, indem neben den klinisch-

- pathologischen Daten - wie erwähnt - zusätzlich neue molekularbiologische Daten berücksichtigt werden. Wie die Figuren 4a und 4b deutlich zeigen, ermöglicht das Verfahren durch den Einsatz eines Data-Mining-Systems, 5 solche Daten bzw. Merkmale - sog. Klassifizierer -, insbesondere molekularbiologische, nicht in den Datensätzen der klinischen Daten enthaltene Merkmale, zu ermitteln, die zu einer Differenzierung der Risikogruppen innerhalb der UICC-Gruppen führen.
- 10 Bei Berücksichtigung eines so ermittelten Merkmals/Klassifizierers, d.h. nach entsprechendem Training des Systems, können Prognosen für diese beiden Untergruppen sehr viel genauer erfolgen. Hier im Beispiel der Fig. 4b ist die Abweichung für die Vorhersage einer 15 Fünf-Jahre-Überlebenszeit für die Patientengruppe ohne dieses Merkmal um 25% nach oben gegenüber der gesamten Gruppe, für die Gruppe der Patienten, welche das Merkmal aufweisen, weicht die Prognose für die Fünf-Jahre-Überlebenszeit um 8% gegenüber der gesamten Gruppe nach 20 unten ab. Eine Prognose wird also bei Verwendung der erfindungsgemäßen Datenauswertung durch die automatische Ermittlung von signifikanten Merkmalen wesentlich präziser.
- 25 Figuren 5 und 6 veranschaulichen detailliert durch verschiedene graphische Darstellungen, wie die Prognosequalität signifikant verbessert werden kann, wenn der Prognose zusätzliche Daten zugrundegelegt werden, die mit Hilfe des Data-Mining Systems der Erfindung ermittelt wurden. 30

Eine weitere genauere Klassifizierung kann erforderlichenfalls innerhalb der UICC-Klassen durchgeführt werden, wenn man Feature-Selection 6 und Trainieren 7

des neuronalen Netzes nur auf eine UICC-Klasse allein anwendet. Die Figuren 5a-c zeigen die Ergebnisse bei der Anwendung auf Patienten des UICC-Stadiums I (Fig. 5a), auf Patienten des UICC-Stadiums II (Fig. 5b) und
5 auf Patienten des UICC-Stadiums III (Fig. 5c). In allen Fällen wird deutlich, daß die Anwendung der erfindungsgemäßen Datenauswertung auf spezielle Patienten-Gruppen eine weitere signifikante Klassifizierung erlaubt.

10 ROC-Kurven veranschaulichen die Güte einer Voraussage. Dabei wird auf der Ordinate die 'Sensitivität' (d.h. das Verhältnis der korrekten Voraussagen über den Eintritt eines Ereignisses zur Gesamtzahl der positiven Testergebnisse) aufgetragen gegen das Komplement der
15 Spezifität (das ist der Anteil von gesunden Personen mit negativem Testergebnis unter allen Personen mit negativem Testergebnis). Die Güte der Prognose wird durch die Fläche unter der Kurve angegeben. Angewendet auf die Vorhersage einer 5-Jahres-Überlebenszeit unter
20 alleiniger Verwendung von prä-operativen (ausschließlich klinisch-pathologischen) Daten 9 und auf Prognosen unter Verwendung von sowohl prä- als auch zusätzlichen, durch die erfindungsgemäße Datenauswertung ermittelten, Daten 10 (insbesondere stellten sich hier post-operati-
25 ven Daten als wichtig heraus) wird durch Figur 6 deutlich gemacht, daß die Berücksichtigung dieser zusätzlichen Daten die Güte der Prognose ganz erheblich verbessert.

Von besonderer Bedeutung ist hierbei, daß durch die
30 Erfindung nur solche Daten in die Auswertung einbezogen werden, welche die Vorhersagequalität tatsächlich verbessern (durch die Feature-Selection ausgewählte Merkmale). Damit wird die riesige Menge von molekularbio-

logischen Daten, die heute bereits zur Verfügung stehen, überhaupt erst behandelbar.

- Im Gegensatz zu klinischen Versuchen, gibt es keine
- 5 Standards für Prognosefaktor-Studien. Fast alle Prognosefaktor-Studien haben leider eher einen erklärenden als einen beweisenden Charakter. Es ist daher wichtig für klinische Forscher, Standards für die Beweise für
- 10 einen Prognosefaktor zu definieren, bevor er in der Praxis angewendet wird. Folgende Richtlinien sollten gelten und liegen dem erfinderischen System zugrunde:
- die Reproduzierbarkeit der Untersuchung im eigenen und in anderen Laboren,
 - Untersuchung ungeachtet des Ergebnisses (assay blinded to outcome),
 - 15 - es sollten von weniger als 15% der Patienten Daten fehlen,
 - einheitliche Behandlung,
 - Hypothese im voraus,
 - 20 - genügend Patienten (> 10 pro Ereignis),
 - durch neue Faktoren wird das derzeitige Wissen durch Vorhersagen ergänzt,
 - angepasste Analysen für die verschiedenen Hypothesen,
 - Studienabgrenzungen müssen vorher spezifiziert
 - 25 werden.

Diese Richtlinien sind nicht nur für Studien sondern auch für den Erfolg eines Prognosefaktors von Bedeutung. Sie wurden in dem erfinderischen System berücksichtigt. Die Anerkennung eines neuen Faktors kann nur

30 erfolgreich sein, wenn mindestens eine belegende Studie vorliegt und wenn die Untersuchungen in mehreren Kliniken reproduzierbar sind. Der Prognosewert sollte über

bisherige Standard-Prognosefaktoren hinaus gehen, und er muß Auswirkungen auf die Therapie haben.

Um eine Prognose zu bestimmen, sollte man von drei
5 Prognosefaktoren ausgehen.

- Tumorbezogene Faktoren: charakterisieren die Krankheit,
- patientenspezifische Faktoren: beziehen sich auf den Patienten,
- 10 - umweltspezifische Faktoren, die sich weder direkt auf den Patienten noch auf den Tumor beziehen.

Dabei sollten vorteilhafterweise die folgenden Punkte berücksichtigt werden, wobei diese erforderlichenfalls
15 gemäß neuen Erkenntnissen ergänzt werden können.

Tumorspezifische Faktoren

Diese Faktoren sind eigentlich immer die Determinanten, die bestimmenden Faktoren für das Ergebnis in
20 Krebspatienten. Die wichtigsten tumorspezifischen Faktoren beziehen sich auf histologische Angaben (Typ, Merkmale) und die anatomische Ausbreitung der Krankheit.

25 - Pathologie des Tumors

Die Tumorpathologie ist entscheidend für die Prognose bei Krebs. Der histologische Typ definiert zwar die Krankheit, aber andere Faktoren, wie z. B. das Stadium oder der Befall von Lymphknoten beeinflussen das Ergebnis.
30

- Ausbreitung der Krankheit

Die anatomische Ausbreitung des Tumors wird gewöhnlich nach den Kriterien der TNM-Klassifikation zu

Größe, Infiltration des Primärtumors, vorhandenen Lymphknotenmetastasen und Fernmetastasen beschrieben.

5 - Tumorbiologie

Bisher wurden krebsspezifische Proteine lediglich als Tumormarker verwendet, um die Tumorbelastung wiederzuspiegeln, ohne jedoch das Tumorverhalten genau charakterisieren zu können. Jüngere Ergebnisse in der Tumorbiologie haben die prognostische Rolle von tumorspezifischen Proteinen wieder in den Vordergrund rücken lassen. Als Genprodukte können sie u. a. Ursachen und Suppression von Krebs, die normale und anormale Kontrolle des Zellzyklus und Metastasierung und Angiogenese des Tumors bestimmen. Neue Technologien in der molekularen Diagnostik machen es heute möglich, genetische Information bezogen auf minimale Tumorbelastung, aggressives Tumorzellwachstum und Tumorzellreaktion infolge von Veränderungen der DNA oder Immuntherapien zu bestimmen.

- Tumorspezifische Symptome

Obwohl sie auch als patientenspezifisch angesehen werden können, ist die eigentliche Ursache für Symptome in der Onkologie der invasive Charakter des Tumors. In der Tat sind Symptome bei den meisten Krebspatienten ein sehr wichtiger Prognosefaktor. Klassische Beispiele für die Wirkung von Symptomen sind die B-Symptome (Nachtschweiß, Fieber und Gewichtsverlust).

Patientenspezifische Faktoren

Dies sind im Patienten vorhandene Faktoren, die entweder indirekt oder gar nicht maligne sind, die jedoch durch eine Interferenz mit dem Tumorverhalten oder ihre Reaktion auf die Behandlung einen großen
 5 Einfluß auf das Ergebnis haben können. Hier unterscheidet man zwischen demografischen Faktoren, Komorbidität und zugleich existierenden Krankheiten.

- Demografische Faktoren

10 Diese Faktoren, die sich auf das onkologische Ergebnis auswirken, sind Alter, Geschlecht und ethnische Zugehörigkeit. Keiner dieser Faktoren kann durch einen Eingriff bzw. eine Behandlung beeinflusst werden, aber viele beeinflussen unabhängig voneinander das
 15 Ergebnis anderer Faktoren. Beispielsweise haben ältere Patienten eine niedrigere Überlebenszeit bei der Hodgkin'schen Krankheit oder bei Lymphomen. Die Rolle des Geschlechts ist weitaus weniger genau definiert, aber bei Hodgkin Erkrankungen oder bösartigen Malignomas waren die Ergebnisse bei Männern
 20 schlechter als bei Frauen.

- Komorbidität

25 Diese Faktoren können vererbte genetische Krankheiten, wie z. B. Neurofibromatose sein, die ein Risikofaktor für neurogene Sarkomas und einen prognostischen Faktor für Krebsergebnisse darstellen.

- Performance Status

30 Der Performance Status ist ein starker prognostischer Faktor für viele Krebsarten, besonders bei denen im fortgeschrittenen Status wie z. B. Lungen- und Blasenkrebs, die eine Chemotherapie erfordern. Als Folge des Alters oder der Komorbidität sollte

dieser Faktor als patientenspezifischer Faktor angesehen werden.

- Übereinstimmung

5 Übereinstimmung mit einer vorgeschlagenen Krebsvorsorgeuntersuchung oder Behandlungsplan kann das Überleben eines Patienten oder einer Gruppe von Patienten beeinflussen. Fehlende Übereinstimmung mit
 10 Krebsvorsorgeempfehlungen bei Brustkrebs kann zu einer verspäteten Diagnose, einem weiter fortgeschrittenen Stadium bei der Diagnose und schlechterem Überleben führen.

Umweltspezifische Faktoren

15 Obwohl die umweltspezifischen Faktoren weniger untersucht wurden und oft nicht in die Diskussionen einbezogen werden, haben sie einen Einfluß auf das Ergebnis für einen einzelnen oder eine ganze Gruppe von Patienten.

20

- Mediziner

Der Behandlungsplan hat eine tiefgreifende Wirkung auf das Ergebnis. Unangemessene Eingriffe können in
 25 exzessiver Toxizität und eingeschränkter Lebensqualität enden. Fehlgeschlagene Kontrolle des Krebses kann auch den Tod für den Patienten bedeuten. Die Expertise des behandelnden Arztes ist ein weiterer prognostischer Faktor, da sie ebenfalls das Ergebnis im Krebspatienten beeinflusst. Es gibt zunehmende Beweise dafür, daß Kliniken, die keine
 30 bestimmte Anzahl („critical mass“) an Patienten behandelt, auch keine optimalen Behandlungsergebnisse erreichen.

- Gesundheitswesen

Hier gibt es große Divergenzen zwischen einzelnen Bevölkerungsgruppen. Mehrere Studien haben belegt, daß z. B. ältere Menschen (75 Jahre und älter) oder Patienten aus anderen ethnischen Gruppen nicht die gleiche Behandlung bekommen wie z. B. jüngere oder einheimische Patienten und somit ihr Behandlungsergebnis beeinflusst wird.

10 - Gesellschaft

Studien des Office of National Statistics (GB) haben gezeigt, daß das Überleben bei einer Krebserkrankung abhängig ist vom sozial-wirtschaftlichen Stand der Patienten. Ein weiterer Faktor für eine schlechtere Prognose ist die Ernährung.

Für die Genexpressionsprofilierung auf RNA- und Protein-Ebene werden purifizierte Proben und klinische Daten von kolorektalen Patienten, von denen die erforderlichen Angaben zur Verfügung stehen, analysiert und daraus die Transkriptom- und Proteom-Profile bestimmt. Hierzu steht eine Vielzahl von tiefgekühlten Proben von verschiedenen kolorektalen Patienten aus verschiedenen Institutionen zu Verfügung.

25

Diese Proben wurden speziell für diesen Zweck entnommen, und sind charakterisiert durch eine ausgezeichnete Qualität und Reproduzierbarkeit. Der Anteil an Epithelzellen variiert erheblich zwischen verschiedenen Präparaten (Reymond et al, Electrophoresis 1997a). Für die Probenvorbereitung wurde eine Methode entwickelt, die es erlaubt, reine Epithelzellen in genügender Menge (über 10^8 Zellen) aus chirurgischen Präparaten vorzubereiten (diese Methode ist z. B. beschrieben in UK Patent Application GB 9705949.7). Aus

35

diesen Proben konnten sowohl Proteine, als auch RNA vorbereitet werden, die qualitativ mit Produkten aus Zelllinien verglichen werden können.

- 5 Die nach dieser Methode vorbereiteten Proben können verglichen werden, auch wenn sie aus verschiedenen Institutionen kommen. Damit ist eine Grundbedingung erfüllt, um später die prädiktiven Aussagen des hier beispielhaft beschriebenen Systems in verschiedenen
- 10 Institutionen zu vergleichen. Damit können in der Zukunft die großen Probendurchsatzzahlen erreicht werden, die zur Validierung der Genexpressionsforschung notwendig sind. Bis heute wurden mehrere Tausend Proben von Patienten, z. T. mit Stuhl, Blut und Knochenmark-
- 15 punktion, gewonnen. Der theoretische Input des klinischen Netzwerkes, in dem das beispielhafte Informationssystem integriert ist, beträgt mehrere tausend neue kolorektale Karzinome pro Jahr.
- 20 Diese Proben sind mit klinisch-pathologischen Daten verbunden, die an die Anforderungen des beispielhaften medizinischen Informationssystems angepaßt sind, d. h. daß alle Parameter (Variablen) in jeder Institution gesammelt worden sind. Das (gemeinsame) Follow-up Schema
- 25 entspricht den deutschen Richtlinien. Die persönlichen Patientendaten werden in den teilnehmenden Institutionen behalten, die das Follow-up der Patienten sichern. Nur anonymisierte Daten werden an das Informationssystem weitergegeben.
- 30 Zur Durchführung einer qualitativ hochwertigen Genexpressionsanalyse steht inzwischen vielfältige Technik zur Verfügung, mit der man in der Lage ist, das Expressionsniveau jedes bekannten Gens auf Transkriptom- und
- 35 Proteomebene zu analysieren. Ein Komplettsystem für

derartige Analysen umfaßt beispielsweise die folgenden Komponenten:

Transkriptom-Analyse

- 5 - Die DNA-Chips (DNA-Mikroarrays): Man unterscheidet grundsätzlich die cDNA-Chips von den Oligo-Chips. Bei cDNA-Chips werden ca. 300 - 400 bp lange PCR Produkte auf die Chips aufgebracht. Man ist derzeit in der Lage ca. 14.000 cDNAs auf ein Array zu
10 spotten.
Bei den Oligo-Chips werden ca. 60 bp lange Oligonukleotide auf der Chip-Oberfläche synthetisiert. Man produziert Arrays mit 8.400 Features, auf Wunsch auch als Doppellarray (16.800 Spots). Gerade im Bereich der DNA-Chips werden Neuentwicklungen in sehr
15 kurzer Zeit zu immer dichteren Arrays (höhere Spotanzahl) mit einer sehr großen Flexibilität bei der Sequenzauswahl führen.
- 20 - Der Mikroarray-Scanner: Neuentwicklungen auf dem Gebiet der Mikroarray-Scanner sind in der Lage gleichzeitig 2 Fluoreszenzwellenlängen zu analysieren. Bei einer Auflösung von 5 oder 10 μm (vom Benutzer einstellbar) benötigt der Scanner ca. 8 min
25 für das scannen eines Chips. Ein 48 Chips fassendes Karussell erlaubt die Benutzung dieses Systems in der Hochdurchsatz-Analyse.
- Der Bioanalytiker: Der Bioanalytiker ist ein Lab-on-a-Chip-System, das zur Qualitätskontrolle vor allem
30 bei der RNA-Aufreinigung dient. Mit Hilfe des Bioanalytikers wird die vom Experimentator aufgereinigte RNA maschinell qualitativ und quantitativ analysiert.

Proteom-Analyse

Mit Proteom-Techniken kann die qualitative und quantitative Expression von Proteinen in verschiedenen Krankheitsstadien bestimmt werden. Da bekannterweise
5 post-translationelle Proteinänderungen eine wichtige Rolle im klinischen Verhalten kranker Zellen, Gewebe und/oder Organe bedeuten, haben diese Unterschiede in der Proteinexpression einen wichtigen Einfluß in der Anwendung des beispielhaft beschriebenen Informations-
10 systems.

Als Proteom-Techniken im menschlichen kolorektalen Karzinom können z. B. ein- (SDS-PAGE) oder zweidimensionale Gel-Elektrophorese (2D PAGE), N-terminale
15 Sequenzierung und Massenspektrometrie (MALDI-TOF und MS-MS) sowie Chips, auf denen Antikörper-, Ligand- oder verschiedene Oberflächen zur Bindung von Proteinen angewendet werden, dienen.

20 Als beispielhafte Technologie für Arbeiten auf dem Gebiet der Proteom-Forschung bietet sich eine Kombination der beiden Schlüsseltechnologien 2-D-Gelelektrophorese und Massenspektrometrie an. Die Proteine werden mittels der 2-D-Gelelektrophorese aufgetrennt
25 und anschließend gefärbt. Die Protein-Spots werden ausgeschnitten, enzymatisch verdaut und das entstehende Peptidgemisch massenspektrometrisch untersucht. Über einen Datenbankabgleich der resultierenden Peptide Mass Fingerprints erfolgt die Proteinidentifizierung.

30 Die Massenspektrometrieplattform setzt sich in diesem Falle aus einer automatischen Probenvorbereitungstation, einem High-Performance MALDI-Massenspektrometer (Matrix-Assisted Laser Desorption/Ionization Time of Flight) und einer Datenstation zur automatischen

Durchführung der Datenbanksuche zusammen. Das MALDI-MS verfügt über eine hohe Sensitivität und hohe Massengenauigkeit; beides sind Grundvoraussetzungen für eine erfolgreiche Proteinidentifizierung. Zusätzlich können
5 mittels der PSD-Technik (Post Sorce Decay) Sequenzinformationen der Peptide ermittelt werden.

Um einen leichteren Zugang zu Sequenzinformationen zu bekommen und gezielt posttranslationale Modifikationen bestimmen zu können, ist der Einsatz eines Elektrospray-Massenspektrometers hilfreich. Durch den Einsatz
10 eines automatischen Spotpickers und eines Digesters kann eine weitere Automatisierung erreicht werden.

Eine weitere wichtige Aufgabe, die bei der Entwicklung
15 des erfindungsgemäßen medizinischen Informationssystems gelöst werden mußte, ist die Übersetzung der verschiedenen Informations-Plattformen (Transkriptom- und/oder Proteomdaten) in eine gemeinsame Sprache.

20 Um dieses Konvertierungsproblem zu lösen, wurde für das beispielhafte onkologische Informationssystem ein Bioinformatik-Konzept entwickelt, daß es erlaubt, Daten aus der Klinik, aus der Pathologie, aus DNA-Datenbanken (wie z. B. CGAP), aus cDNA-Arrays (wie z. B. Agilent
25 Chips) und aus 2D PAGE zu integrieren und zu analysieren. Die verschiedenen Informationen aus der Klinik, der Pathologie, der Transkriptom- und der Proteomforschung werden in die Web-basierte (*.xml) bioinformatische Sprache GEML (Gene Expression Markup Language)
30 übersetzt (s. <http://www.geml.org>).

Damit wurden die Voraussetzungen für ein echtes "Bridging" der verschiedenen Informationsbasen geschaffen. Dieses „Bridging“ ist als eine unabdingbare Vor-

aussetzung für die Umsetzung der experimentellen Ergebnisse der Gen-Expression in eine klinisch sinnvolle Information zu betrachten.

5 Nachdem nun die Daten aus Klinik und Labor in das *.xml Format übersetzt wurden, kann man auf eine (vereinheitlichte) Datenbank zugreifen, die etwa 10^4 Informationen enthält. Diese etwa 10^4 Informationen, die jetzt für jeden Patienten zur Verfügung stehen, können
10 keinesfalls dem Onkologen direkt vermittelt werden. Aus diesem Grunde muß in das erfindungsgemäße Informationssystem eine Software integriert sein, die eine Auswertung dieser Fülle an Information unterstützt.

15 Als Komponente des beispielhaften onkologischen Informationssystems findet Verfahren zur Datenreduktion Verwendung. Dabei mußte verfügbare Datenreduktions-Software an die speziellen Anforderungen der onkologischen Auswertungen angepaßt werden. Die etwa 10^4 Informationen pro Patient werden durch den Einsatz dieser Software
20 auf 10^2 reduziert.

Die vom Scanner erzeugten digitalisierten Proteom- oder Transcriptom-Bilder werden in einem kompatiblen
25 Analyseprogramm verarbeitet. Dieses Programm ist in der Lage die Genexpressionsdaten auszuwerten und zu speichern. Das Programm protokolliert jedes Genexpressionsmuster und erlaubt Vergleiche verschiedener Experimente. Hierzu sind i. a. Datenbankabfragen in externen
30 sowie internen Datenbanken notwendig. Das Programm generiert technologie-spezifische Irrtum-Modelle. Die aus den Irrtum-Modellen generierten Probabilitätswerte jeder Messung werden über die gesamte Analyse-Umgebung propagiert, was einen höheren prädikativen Wert bei

Cluster-Analysen, Ähnlichkeiten-Suchen und Tendenz-Analysen ermöglicht. Durch besondere informationstechnische Werkzeuge ermöglicht das Programm, Analysen über Exons, Sequenzen, Cluster Intensität- und Verhältnisrechnungen durchzuführen. Clustering-Analysen beinhalten beispielweise agglomerative, Teilungs-, Mittelwerte und Medianwerte-Algorithmen. Ein beispielhaftes informationstechnisches Verfahren erlaubt es, ähnliche Muster mit dem Muster von Interesse innerhalb aller Datensets der Datenbank zu recherchieren. Auch Zeit-Sequenzen, beispielhaft in einer iterativen Nachsorge-messung, können über eine Zeitlinie dargestellt werden, wodurch spezifische Verhalten identifiziert werden können. Besondere Suchmaschinen erlauben eine rasche Datenbank-anfrage, und können an eine interne Datenbank angepaßt werden. Auch Hypertext-Links können formuliert werden, so daß Verbindungen mit internen oder externen Datenbanken hergestellt werden können.

20 Durch dieses Vorgehen wird erreicht, daß nur die biologischen Daten, die für eine bestimmte klinische Beobachtung ein signifikantes Verhalten zeigen, berücksichtigt und für die Auswertung übernommen werden. Dadurch wird gegenüber herkömmlichen medizinischen Informationssystemen eine erhebliche Datenreduktion erreicht.

Die Bioinformatik wird weiterhin dadurch unterstützt, daß klinische Outcomes (wie die Metastasierungsfähigkeit oder die Therapie-Resistenz eines bestimmten Tumors) nach Berücksichtigung der klinisch-pathologischen Daten direkt mit Datenmustern verbunden werden können. Diese Interpretation kann z. B. durch künstliche Intelligenz und/oder maschinelle Lernverfahren

vereinfacht werden. Herkömmliche Computerprogramme umfassen eine Menge von expliziten Anweisungen, die dem Programm genau sagen, was und wie es eine Berechnung auszuführen hat. Systeme der künstlichen Intelligenz (KI-Systeme) arbeiten unter vollkommen anderen Voraussetzungen: dem Programm werden eher Kenntnisse vermittelt, als daß ihm exakte Anweisungen für die Abarbeitung gegeben werden. Dies passiert während der Trainingsphase des KI-Systems. Indem das KI-System wiederholt auf historische Daten angewandt wird und die Ergebnisse dieser Auswertungen ("Schlußfolgerungen") mit den tatsächlich vorliegenden Fakten verglichen werden, lernt es im Verlaufe dieses Trainings das Verhalten, das von dem "fertig ausgebildeten" System gefordert wird.

Der Korrespondenzanalyseansatz und die aufsteigend hierarchische Klassifikation, die bei dem erfindungsgemäßen Informationssystem angewendet werden, weichen signifikant von dem eher klassischen Ansatz der Diskriminanzanalyse mittels Hauptkomponentenanalyse ab. Beginnend mit einer Reihe von Experimenten, wobei jedes Experiment eine große Anzahl an Datenpunkten aufweist, liefert die Korrespondenzanalyse für die Repräsentation der Proben einen faktorisierten Raum von reduzierter Größe. Die ansteigende hierarchische Klassifikation sortiert die Bilder in aussagefähige Gruppen. Die simultane Darstellung sowohl der Spots als auch der Chip- oder Gel-Bilder findet in dem selben faktorisierten Raum statt. Die charakteristischen Gen- oder Protein-Vertreter einer bestimmten Klasse von Gelen (z. B. Krebsmetastasen-Proben) werden präzise markiert, was die Analyse erheblich vereinfacht. Dementsprechend kann die Software automatisch Protein- oder Gen-Muster

klassifizieren, wobei dafür den jeweiligen Anforderungen entsprechend Hauptkomponentenanalyse, Wavelet-Analyse, künstliche neuronale Netze, heuristischer Clusterbildungsanalyse und andere einzeln oder in
5 Kombination eingesetzt werden können.

Um die mit der Software ausgewählten Parameter flächen-
deckend und billig analysieren zu können, wird ein
reproduzierbarer Labortest durchgeführt.

10

Dieser Labortest erlaubt es, Outcome-relevante genetische, translationelle oder funktionelle Charakteristiken eines Tumors zusammenzufassen. Dieses Vorgehen ermöglicht es, sogenannte Integrated Health-Care
15 Solutions anzuwenden, wo die Therapie mit der Diagnostik gekoppelt ist.

Ein solcher Labortest (oder auch mehrere Labortests) kann (bzw. können) beispielhaft mit einem Chip durchgeführt werden. Die Chips weisen folgende Eigenschaften
20 auf:

- der Chip liefert Daten zum Metastasierungspotential,
- der Chip liefert Daten zur Therapieantwort bei mindestens 10 verbreiteten Chemotherapeutika,
- 25 - der Chip liefert Daten zum Patientenmetabolismus (z. B. enzymatische Ausrüstung),
- der Chip liefert Information zur Autoimmunität gegen das Tumor,
- der Chip beinhaltet nicht mehr als 10^2 verschiedene
30 Informationen (+ doubles) und
- der Chip ist reproduzierbar.

Durch die Reproduzierbarkeit wird eine breite Anwendbarkeit der Chips erreicht, so daß eine preiswerte Herstellung möglich wird.

- 5 Da die relevanten biologische Daten sich je nach Diagnose unterscheiden, muß für jede Diagnose ein eigener Test entwickelt werden. Die daraus resultierenden Labortests können sich dabei erheblich von denen unterscheiden, die beim kolorektalen Karzinom Outcome-
10 relevant sind. Dadurch ist es schwierig, dieses Labor-test genauer zu beschreiben.

Für den Datenaustausch zwischen behandelndem Arzt, Referenzlabor(en) und Datenbanken umfaßt das erfindungsgemäße Informationssystem eine vorzugsweise mehrsprachige, sichere Web-Schnittstellen, die bei der beispielhaften Lösung die Verbindung mit Onkologen und Referenzlabors ermöglichen. Mit der Strukturierung des Informationssystems mittels des Internetstandards XML
15 (eXtensible Markup Language) verbessert sich die Verfügbarkeit klinisch relevanten Wissens zu einem spezifischen medizinischen Problem am Arbeitsplatz des Onkologen.
20

- 25 Die gewählte kryptographische Basistechnik des beispielhaften onkologischen Informationssystems ist die symmetrische Verschlüsselung. Hier stehen hocheffiziente Verfahren zur Verfügung, die bei einer Schlüssellänge von beispielweise 128 bits langfristige Sicherheit
30 gewähren. Kommunikationspartner verfügen über einen gemeinsamen Schlüssel, die PIN-Nummer. Die PIN-Nummer wird nur dem behandelten Arzt gegen Rechnung zugestellt, so daß der Patient keinen direkten Zugang zu den Aussagen des Informationssystems bekommen kann.

Als vorteilhafter Standard kann beispielweise der AES (Advanced Encryption Standard) gewählt werden.

5 Als sicherer Aufbewahrungsort für die elektronische Identität eines bestimmten Onkologen ist die kryptographische Chipkarte ideal, also im Gesundheitswesen die HPC (Health Professional Card, Ausweis für Berufe im Gesundheitswesen).

10 Zunächst wird eine Plattform definiert. Das heißt Inputs und Outputs werden festgelegt. Hierzu werden Festlegungen zum Informationsfluß und zum Abrechnungsmodell getroffen.

15 Für diesen Informationsfluß wird beispielhaft der HON Code of Conduct (HONcode) für medizinische Webseiten im Gesundheitsbereich umgesetzt (www.hon.ch/HONcode/German). Auf aktive Inhalte wie Java-Skripte wird verzichtet, außer in zwingenden Anwendungsfällen wie der Fern-Eingabe von mit PIN anonymisierten klinischen und pathologischen Daten und von Nachsorgedaten.

20

Konkrete Maßnahmen sind auch für die Sicherheit des Servers des beispielhaften Informationssystem getroffen worden. Nur die nötigsten TCP-Dienste laufen auf dem Server. Der Mail-Server wird mit aktuellen Virusfiltern ausgestattet. Client/ Server-Verbindungen, auch mit anonymisierten Daten, werden über SSL (Secure Socket Layer) abgewickelt. Für die Nutzer-Authentisierung werden beispielhaft X.509 Zertifikate eingesetzt. Schließlich wird die Datenbank des beispielhaften Informationssystems regelmäßig gesichert, um nach einem Schadenfall wieder einen sauberen Zustand herstellen zu können.

25

30

Der Betreiber dieses Informationssystems kann zur kommerziellen Nutzung der Erfindung entweder selbst einen Vertrieb aufbauen oder beispielsweise Referenzlabors,
5 große Pharma-Unternehmen oder Internet Content Provider mit dem Vertrieb beauftragen. Diese wiederum könnten das Informationssystem dem Zielpublikum zur Nutzung anbieten. Die Referenzlabors, großen Pharma-Unternehmen oder Content Provider stellen dabei die eigentlichen
10 Kunden des Betreibers („Kunden“) dar; das Zielpublikum bzw. Nutzer des Informationssystems („Teilnehmer“) sind z. B. Ärzte, die sich mit Krebserkrankungen beschäftigen.

Dabei würden durch diese Geschäftsidee mehrere Vorteile
15 genutzt. So müßte sich der Betreiber nur auf wenige Kunden konzentrieren und könnte z. B. das entwickelte Vertriebssystem dieser großen Kunden nutzen, das bei den großen Pharma-Unternehmen bis zu dem einzelnen Arzt reicht. Durch die geeignete Wahl der Kunden kann die
20 weltweite Verfügbarkeit des Informationssystems erreicht werden.

Die Abrechnung der Gebühren für die Nutzung des Informationssystems sollte vorteilhafterweise ebenfalls über die Kunden und nicht direkt über die einzelnen Teilnehmer aus dem Zielpublikum erfolgen. Dabei könnte man
25 je nach Erfordernis eine Vorauszahlung, Ratenzahlung, Gebühren pro Nutzungsfall oder Gebühren prozentual zum Umsatz, den der Kunde mit dem Informationssystem macht, vereinbaren. Auch Gebühren je „PIN“, die an die Teilnehmer des Zielpublikums oder Kunden vergeben wurden,
30 sind denkbar. Diese PIN ermöglicht dem Teilnehmer aus dem Zielpublikum den Zugang zum Informationssystem und dessen Nutzung; genauer: zu den mit der PIN verknüpften patientenspezifischen Daten.

Der Teilnehmer erhält die PIN gegen Bezahlung einer Gebühr an den Kunden („no money - no PIN“). In Verbindung mit der PIN erhält der Teilnehmer einen Chip, auf welchem die für die Analyse erforderlichen Tests enthalten sind. Die Anforderungen an den Chip (Art und Anzahl der enthaltenen Tests) ergeben sich u. a. auch aus den Aussagen, die das Informationssystem hinsichtlich der Signifikanz der Variablen macht. Nach Aufbringung der Patientenprobe wird der Chip dann an ein Referenzlabor geschickt, und die Auswertung erfolgt in der oben beschriebenen Weise. Ebenso ist es denkbar, daß die Patientenprobe vom Teilnehmer direkt an das Referenzlabor geschickt wird und erst dort auf den Chip aufgebracht werden. Die PIN würde dann vom Referenzlabor gemeinsam mit den Testergebnissen dem Teilnehmer übersandt. Der Preis für die Nutzung des Informationssystems wäre im Gesamtpreis für den Kauf des Chips enthalten.

Wenn der Teilnehmer aus dem Zielpublikum eine Anfrage an das Informationssystem stellt, wird er in der Regel aufgefordert, gewisse Angaben zu machen, insbesondere über Verlauf der Therapie, Medikamentierung oder Verlauf der Erkrankung. Diese Daten dienen u. a. dazu, das System zu optimieren. Da damit der Wert des Systems gesteigert wird, kann man von dieser Dateneingabe möglicherweise auch eine Rückerstattung eines gewissen Betrages der Gebühren an den jeweiligen Teilnehmer erwägen.

Die Erfindung ist nicht beschränkt auf die hier dargestellten Ausführungsbeispiele. Vielmehr ist es möglich, durch Kombination und Modifikation der genannten Mittel und Merkmale weitere Ausführungsvarianten zu realisieren, ohne den Rahmen der Erfindung zu verlassen.

Bezugszeichenliste

	1	Input-Variablen
5	2	Modul zur Berechnung der Korrelation
	3	Modul zur multivarianten statistischen Analyse
	4a	Resultate des herkömmlichen Verfahrens
	4a	Resultate des erfindungsgemäßen Verfahrens
	5	Transformationsschritt
10	6	Feature-Selection
	7	Training und Auswahl des Modells
	8	bestes Modell, Modellarchitektur
	9	prä-operative Daten
10	10	prä-operative und zusätzliche, durch die erfindungsgemäße Datenauswertung ermittelten, Daten
15		

Patentansprüche

- 5 1. Verfahren zur Datenauswertung unter Verwendung
von mit Datenbanken gekoppelten Datenverarbei-
tungseinrichtungen,
dadurch gekennzeichnet, daß
- 10 der Datenverarbeitungseinrichtung zugeführte An-
fragedaten ausgewertet werden, indem nach vorgeb-
baren Regeln und/oder mit Verfahren der künst-
lichen Intelligenz mit den Anfragedaten korres-
pondierende, in der (den) Datenbank(en) gespei-
cherte Daten ermittelt werden,
- 15 die Güte dieser korrespondierenden Daten automa-
tisch bewertet,
anhand dieser Bewertung, der zugehörigen Anfrage-
und/oder korrespondierenden Daten automatisch die
Signifikanz dieser Anfrage- und/oder korrespon-
dierenden Daten für die Anfrage ermittelt wird,
- 20 die Ergebnisse der Auswertung, die Gütebewertung
und/oder die Signifikanz der Daten ausgegeben
und/oder abrufbar bereitgestellt werden.
- 25 2. Verfahren nach Anspruch 1,
dadurch gekennzeichnet, daß
- 30 die Datenauswertung medizinische Prognosen gene-
riert und die Anfragedaten als klinisch-patholo-
gische Daten durch einen Arzt und/oder als biomo-
lekulare Daten durch ein Analyselabor der Daten-
verarbeitungseinrichtung zugeführt werden.

3. Verfahren nach einem der Ansprüche 1 oder 2,
dadurch gekennzeichnet, daß
die Anfragedaten und/oder die in der (den) Daten-
bank(en) gespeicherten Daten
- 5 - krankheitsbezogene Faktoren und/oder
- patientenspezifische Faktoren und/oder
- umweltspezifische Faktoren
berücksichtigen.
- 10 4. Verfahren nach einem der Ansprüche 1 bis 3,
dadurch gekennzeichnet, daß
eine Aktualisierung der in der (den) Daten-
bank(en) gespeicherten Daten Erfahrungsdaten er-
folgt, indem Daten über die Therapie und den
15 Krankheitsverlauf zu von der Datenverarbeitungs-
einrichtung prognostizierten Fällen der Daten-
verarbeitungseinrichtung zugeführt werden.
- 20 5. Verfahren nach einem der Ansprüche 1 bis 4,
dadurch gekennzeichnet, daß
eine Aktualisierung der Bewertungsvorschriften in
einem iterativen Lernprozeß erfolgt, in welchem
Anfrage, Anfragedaten, bei der Auswertung der An-
frage benutzte, in der (den) Datenbank(en) ge-
25 speicherten Daten, Ergebnisse der Auswertung und
tatsächlich eingetretene Ereignisse berücksich-
tigt werden.

6. Verfahren nach einem der Ansprüche 1 bis 5,
dadurch gekennzeichnet, daß
in Abhängigkeit von der Bewertung die Menge der
in der (den) Datenbank(en) gespeicherten Daten
5 reduziert wird.
7. Verfahren nach einem der Ansprüche 1 bis 6,
dadurch gekennzeichnet, daß
in Abhängigkeit von der Signifikanz die Zahl der
10 zuzuführenden Anfragedaten und/oder die Menge der
in der (den) Datenbank(en) gespeicherten Daten
reduziert wird.
8. Verfahren nach einem der Ansprüche 1 bis 7,
15 dadurch gekennzeichnet, daß
die Auswertung durch
- Cluster-Analyse und/oder
- Ähnlichkeiten-Suche und/oder
- Tendenz-Analyse und/oder
20 - Korrespondenzanalyse und/oder
- aufsteigend hierarchische Klassifikation und/
oder
- Hauptanalyse und/oder
- Wavelet-Analyse
25 in Verbindung mit aus Irrtums-Modellen generier-
ten Probabilitätswerten und/oder Mitteln der
künstlichen Intelligenz wie
- neuronalen Netzen und/oder
- regelbasierten Systemen
30 erfolgt.

9. Verfahren nach einem der Ansprüche 1 bis 8,
dadurch gekennzeichnet, daß
die Datenauswertung die Ermittlung der Eintritts-
wahrscheinlichkeit von durch die Anfrage und/oder
5 durch die korrespondierenden Daten bezeichneten
Ereignissen umfaßt.
10. Verfahren nach einem der Ansprüche 1 bis 9,
dadurch gekennzeichnet, daß
10 die Signifikanz von Daten derart ermittelt wird,
daß eine Anfrage ein erstes Mal ohne Berück-
sichtigung dieser Daten und ein zweites Mal unter
Berücksichtigung dieser Daten ausgewertet wird,
die Bewertungen der Ergebnisse dieser beiden Aus-
15 wertungen miteinander verglichen und das Maß des
Einflusses der Daten relativ zu einer Verbes-
serung bzw. Verschlechterung der zweiten Bewer-
tung gegenüber ersten Bewertung bestimmt wird.
- 20 11. Verfahren nach Anspruch 10,
dadurch gekennzeichnet, daß
bei einer Verbesserung der zweiten Bewertung ge-
genüber der ersten Bewertung die Daten als signi-
fikant angesehen und bei künftigen Auswertungen
25 berücksichtigt werden,
bei einer Verschlechterung der zweiten Bewertung
gegenüber der ersten Bewertung die Daten als
nicht signifikant angesehen und für künftige Aus-
wertungen nicht mehr genutzt werden.

12. Verfahren nach einem der Ansprüche 1 bis 11,
dadurch gekennzeichnet, daß
die Zuführung der Daten, die Anfrage und/oder die
Ausgabe der Ergebnisse über das Internet erfolgt.
- 5
13. Anordnung mit mindestens einem Prozessor, der
(die) derart eingerichtet ist (sind), daß ein
Verfahren zur Datenauswertung durchführbar ist,
wobei der Datenverarbeitungseinrichtung zugeführte
Anfragedaten ausgewertet werden, indem nach
vorgebbaren Regeln und/oder mit Verfahren der
künstlichen Intelligenz mit den Anfragedaten korrespondierende, in der (den) Datenbank(en) gespeicherte Daten ermittelt werden,
die Güte dieser korrespondierenden Daten automatisch bewertet,
anhand dieser Bewertung, der zugehörigen Anfrage- und/oder korrespondierenden Daten automatisch die Signifikanz dieser Anfrage- und/oder korrespondierenden Daten für die Anfrage ermittelt wird,
die Ergebnisse der Auswertung, die Gütebewertung und/oder die Signifikanz der Daten ausgegeben und/oder abrufbar bereitgestellt werden.
- 10
- 15
- 20
- 25
14. Anordnung nach Anspruch 13,
gekennzeichnet durch
mindestens eine mit mindestens einer Datenbank gekoppelte Datenverarbeitungseinrichtung,
Mittel zur Datenein- und/oder -ausgabe,
nach vorgebbaren Regeln und/oder mit Verfahren der künstlichen Intelligenz arbeitende Mittel zur Ermittlung von in der (den) Datenbank(en) gespeicherten, mit einer Datenverarbeitungseinrichtung
- 30

- zugeführten Anfragedaten korrespondierenden Daten,
Mittel zur automatischen Bewertung der Güte der korrespondierenden Daten,
5 Mittel zur automatischen Ermittlung der Signifikanz der Anfrage- und/oder korrespondierenden Daten für eine Anfrage.
15. 10 Computerprogrammprodukt, das ein computerlesbares Speichermedium umfaßt, auf dem ein Programm gespeichert ist, das es einem Computer ermöglicht, nachdem es in den Speicher des Computers geladen worden ist, ein Verfahren zur Datenauswertung durchzuführen, wobei die Datenauswertung die Ver-
15 fahrensschritte gemäß einem der Ansprüche 1 bis 12 umfaßt.
16. 20 Computerlesbares Speichermedium, auf dem ein Programm gespeichert ist, das es einem Computer ermöglicht, nachdem es in den Speicher des Computers geladen worden ist, ein Verfahren zur Datenauswertung durchzuführen, wobei die Datenauswertung die Verfahrensschritte gemäß einem der Ansprüche 1 bis 12 umfaßt.
25
17. 30 Verwendung eines Verfahrens nach einem der Ansprüche 1 bis 12 zur
- Auswertung von klinischen, pathologischen und/oder molekulargenetischer Daten,
 - Ermittlung der prognostischen Signifikanz von klinischen, pathologischen und/oder molekulargenetischer Daten,
 - Selektion molekularer Targets,

- Schätzung des individuellen Risikos, wie beispielsweise des Metastasierungsrisikos, einzelner Patienten,
 - Schätzung der Probabilität der Therapieantwort z. B. auf Chemotherapeutika und/oder
 - automatischen Generation von Prognose- und/oder Therapievorschlügen.
- 5
18. Daten, Gene, molekulare und/oder genetische
10 Targets, welche durch ein Verfahren nach einem der Ansprüche 1 bis 12, durch eine Anordnung nach einem der Ansprüche 13 oder 14, durch ein Computerprogrammprodukt nach Anspruch 15, durch ein computerlesbares Speichermedium nach Anspruch 16
15 oder eine Verwendung nach Anspruch 17 verfügbar gemacht werden.
19. Herstellungsverfahren für diagnostische Anordnungen, welches die Schritte eines Verfahrens
20 nach einem der Ansprüche 1 bis 12 und einen zusätzlichen Schritt umfaßt, in welchem ein diagnostisch effektives Analysewerkzeug, wie z. B. ein RNA- oder Protein-Chip, und/oder eine Kombination von Genen, welche durch ein Verfahren
25 nach einem der Ansprüche 1 bis 12, durch eine Anordnung nach einem der Ansprüche 13 oder 14, durch ein Computerprogrammprodukt nach Anspruch 15, durch ein computerlesbares Speichermedium nach Anspruch 16 oder eine Verwendung nach Anspruch
30 17 verfügbar gemacht wurden, zusammengestellt wird.
20. Verwendung von Genen oder Kombination von Genen, welche durch ein Verfahren nach einem der An-

- sprüche 1 bis 12, durch eine Anordnung nach einem der Ansprüche 13 oder 14, durch ein Computerprogrammprodukt nach Anspruch 15, durch ein computerlesbares Speichermedium nach Anspruch 16 oder eine Verwendung nach Anspruch 17 verfügbar gemacht wurden, für die Vorbereitung einer diagnostischen Zusammenstellung zur Klassifizierung genetisch bedingter Erkrankungen, Tumoren u.ä. und/oder zur Vorhersage genetisch bedingter Erkrankungen und/oder zur Assoziierung von molekular-genetischen Parametern mit klinischen Parametern und/oder zur Identifizierung von Tumoren durch Genexpressionsprofile.
- 5
- 10
- 15 21. Trägerelement, auf dem Daten, Gene, molekulare und/oder genetische Targets bereitgestellt werden, welche durch ein Verfahren nach einem der Ansprüche 1 bis 12, durch eine Anordnung nach einem der Ansprüche 13 oder 14, durch ein Computerprogrammprodukt nach Anspruch 15, durch ein computerlesbares Speichermedium nach Anspruch 16 oder eine Verwendung nach Anspruch 17 verfügbar gemacht werden.
- 20
- 25 22. Trägerelement nach Anspruch 21, dadurch gekennzeichnet, daß
- das Trägerelement als Chip ausgebildet ist und
- Daten zum individuellen Risiko, wie z. B. Metastasierungspotential, und/oder
 - 30 - Daten zur Therapieantwort z. B. auf Chemotherapeutika und/oder
 - Daten zum Patientenmetabolismus und/oder
 - Information zur Autoimmunität z. B. gegen einen Tumor

bereitstellt.

23. Trägererelement nach Anspruch 22,
dadurch gekennzeichnet, daß
5 die Anzahl der auf dem Trägererelement bereitgestellten Daten Einhundert nicht übersteigt.
24. Trägererelement nach einem der Ansprüche 22 oder
23,
10 dadurch gekennzeichnet, daß
das Trägererelement als reproduzierbarer Chip ausgebildet ist.
25. Methode zur Nutzung eines Systems zur Datenauswertung nach einem der Ansprüche 1 bis 14,
15 dadurch gekennzeichnet, daß
der Zugang zu dem System durch eine gebührenpflichtige PIN ermöglicht wird, wobei die PIN mit
Mitteln zur Erfassung von in das System einzugehenden Daten und/oder mit Mitteln zur Erfassung
20 von Material, welches zur Ermittlung der einzugehenden Daten dient, verbunden ist, und der Nutzer die PIN durch Zahlung einer Gebühr erwirbt.
- 25 26. Methode nach Anspruch 25,
dadurch gekennzeichnet, daß
der Nutzer die PIN zusammen mit dem (den)
Mittel(n) zur Erfassung der Daten und/oder des
Materials beim Kauf dieses (dieser) Mittel(s)
30 erwirbt.

27. Methode nach einem der Ansprüche 25 oder 26,
dadurch gekennzeichnet, daß
eine PIN mit vorgebbaren, in dem System gespeicherten Daten verknüpft ist und die PIN nur
5 Zugang zu diesen mit ihr verknüpften Daten ermöglicht.
28. Methode nach einem der Ansprüche 25 bis 27,
dadurch gekennzeichnet, daß
10 ein Vertreiber des Systems zur Datenauswertung mit mindestens einem Kunden Nutzungsvereinbarungen schließt und der (die) Kunde(n) das System für weitere Teilnehmer durch Vergabe gebührenbehafteter PINs nutzbar macht (machen).
- 15 29. Methode nach einem der Ansprüche 25 bis 28,
dadurch gekennzeichnet, daß
- Referenzlabors und/oder
- Pharmaunternehmen und/oder
20 - Content Provider
das System zur Datenauswertung als Kunde nutzen.
30. Methode nach einem der Ansprüche 25 bis 29,
dadurch gekennzeichnet, daß
25 die Gebühren für die Nutzung des Systems zur Datenauswertung über den Kunden erhoben werden und
- pro Nutzungsfall und/oder
- prozentual zum Umsatz, den der Kunde mit dem
30 System macht und/oder
- pro vergebener PIN
festgelegt werden.

31. Methode nach einem der Ansprüche 25 bis 30,
dadurch gekennzeichnet, daß
ein Mittel zur Erfassung von Material ein Träger-
Chip für die bei einem - erforderlichenfalls
5 reproduzierbaren - Labortest benötigten Proben,
wie z. B. ein DNA-Mikroarray, ist.
32. Methode nach einem der Ansprüche 25 bis 31,
dadurch gekennzeichnet, daß
10 ein Nutzer des Systems zur Datenauswertung die
PIN vom Vertreiber des Systems oder von einem
Kunden des Betreibers erwirbt.
33. Verwendung eines Systems zur Datenauswertung nach
15 einem der Ansprüche 1 bis 32 zur Durchführung von
Profit- oder Non-Profit-Aktionen durch Ärzte,
Patienten und/oder das System betreibende Firmen.
34. Verwendung eines Systems zur Datenauswertung nach
20 Anspruch 33,
dadurch gekennzeichnet, daß
eine Aktion durch
- Teilnehmer und/oder
- Anbieter
25 des Systems gestartet wird.
35. Verwendung eines Systems zur Datenauswertung nach
einem der Ansprüche 33 oder 34,
dadurch gekennzeichnet, daß
30 eine Aktion
- die Schaffung, Unterhaltung und/oder Vermark-
tung eines Exzellenz-Netzwerkes und/oder

- die Distribution von Therapien und/oder
 - die Selektion von Patientengruppen für klinische Studien
- umfaßt.

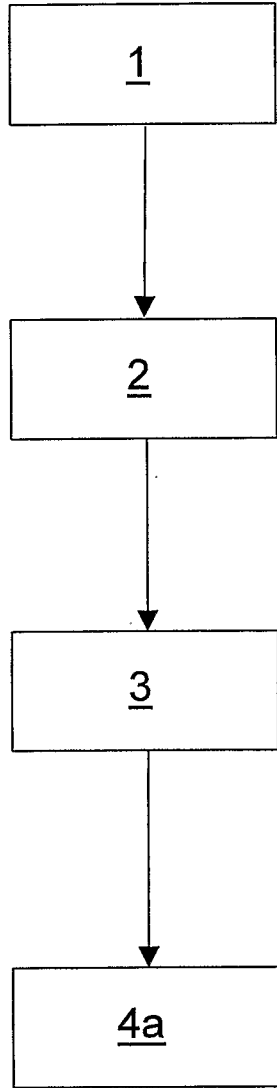


Fig 1a

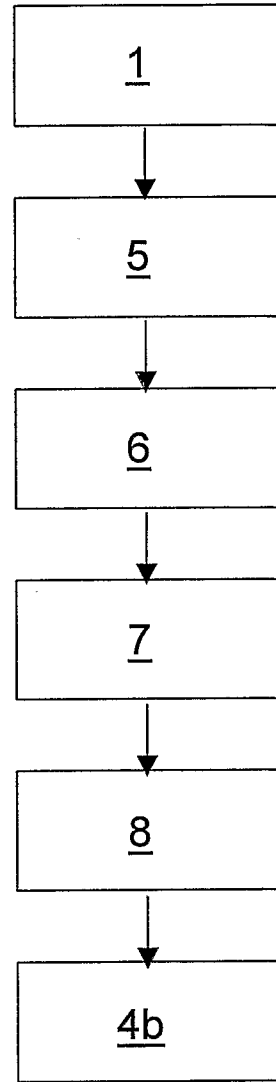


Fig 1b

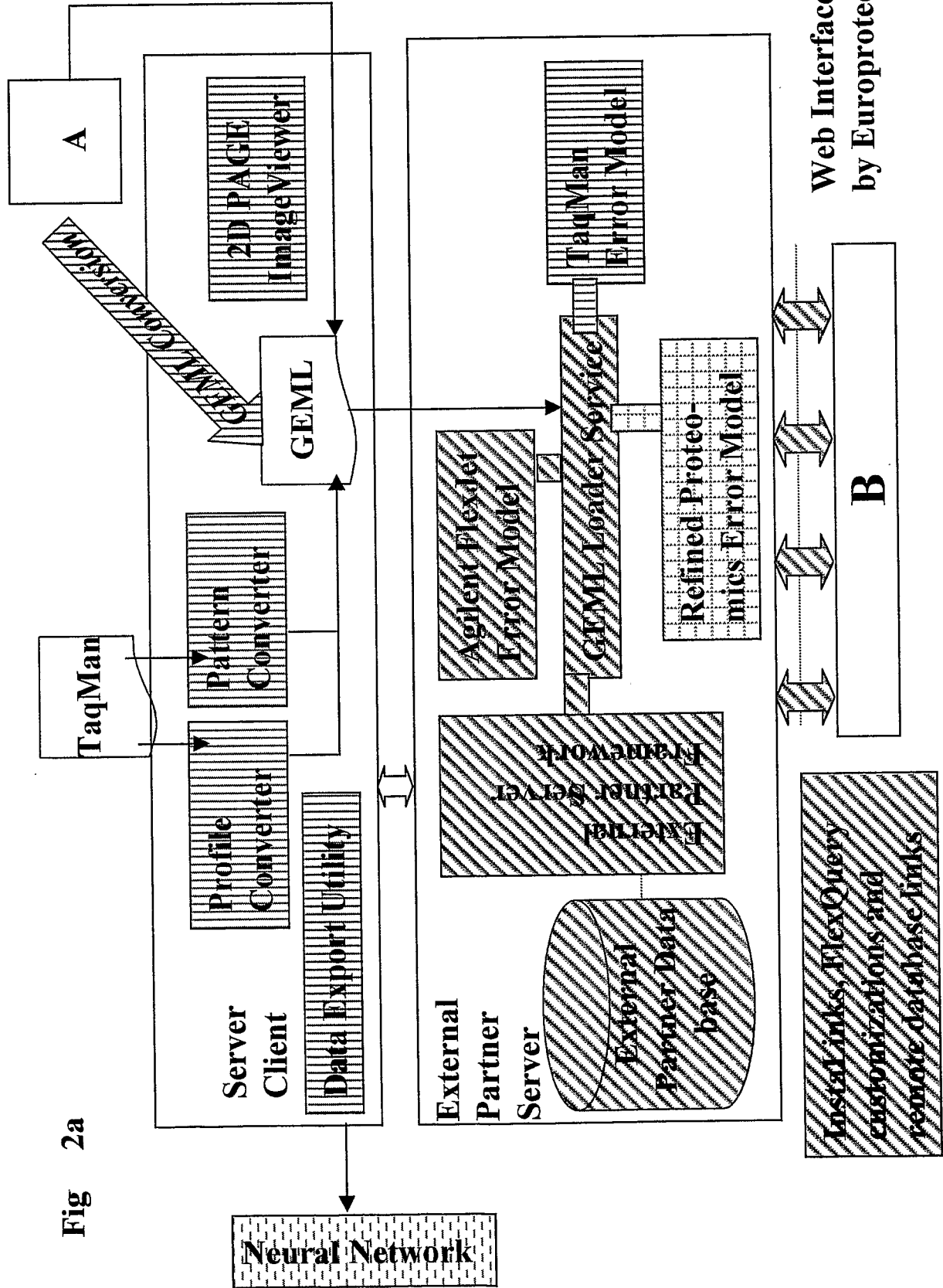


Fig 2a

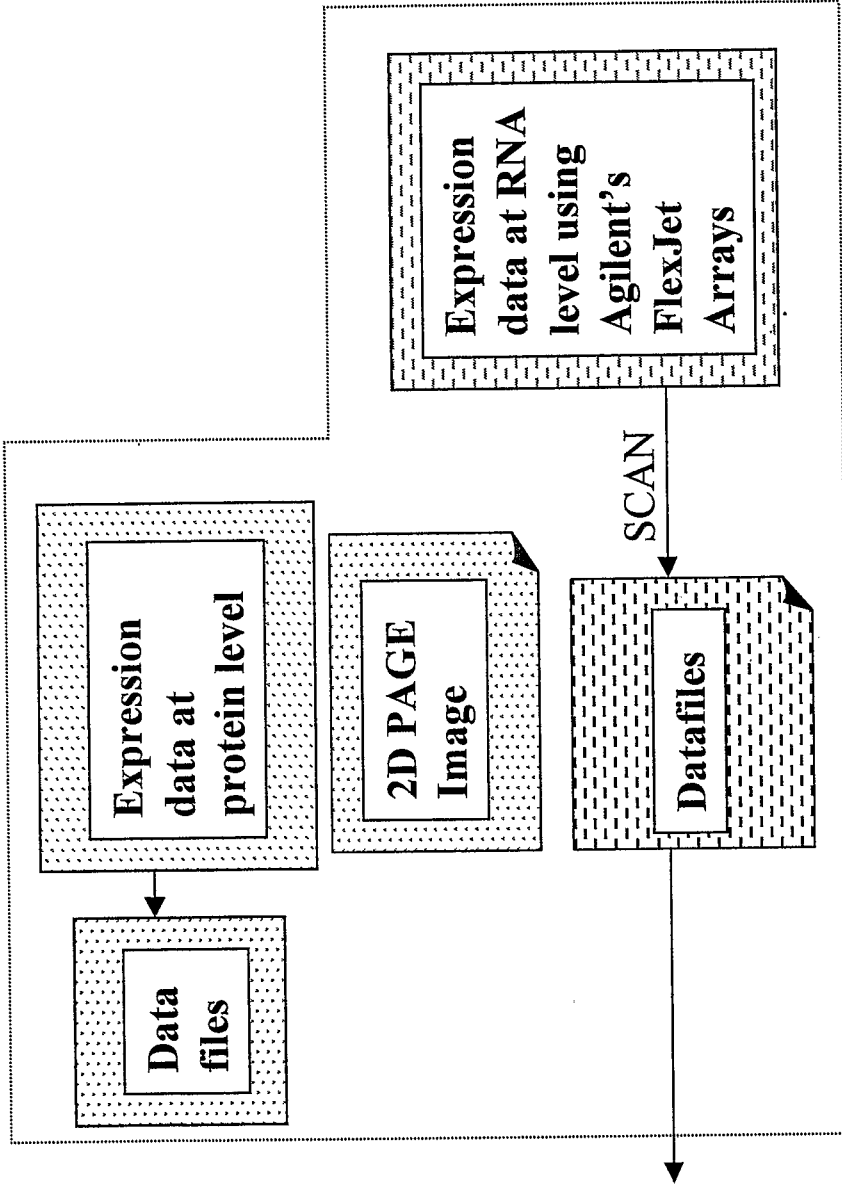


Fig 2b

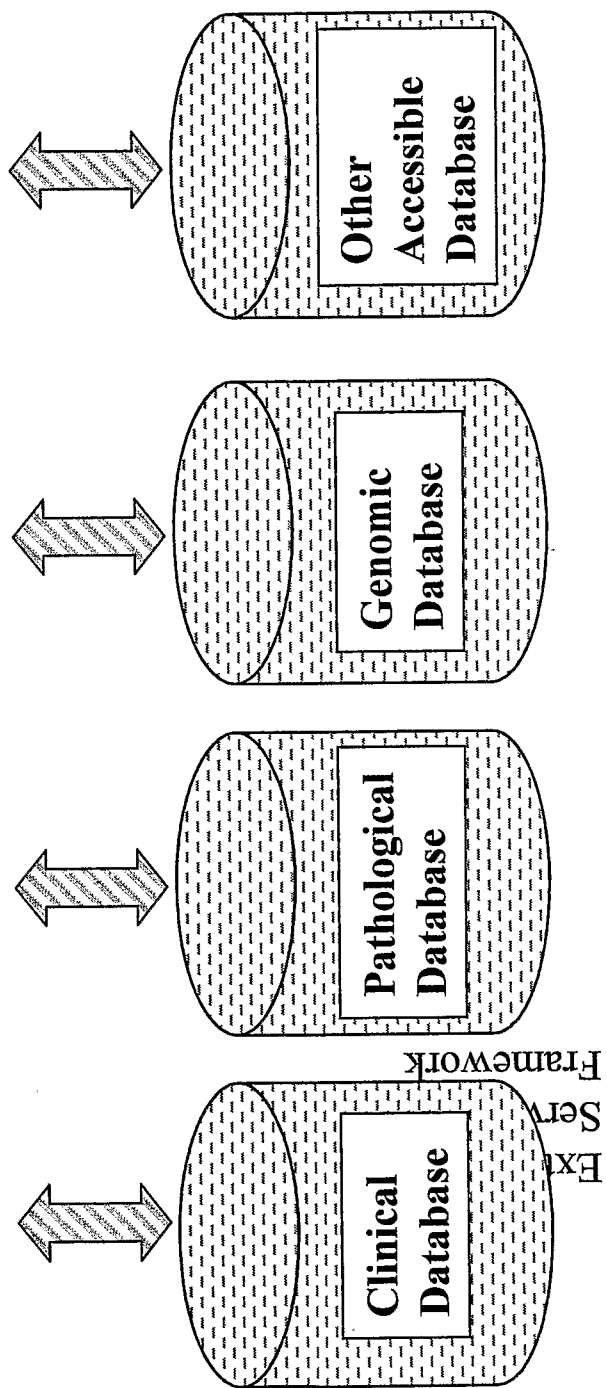


Fig 2c

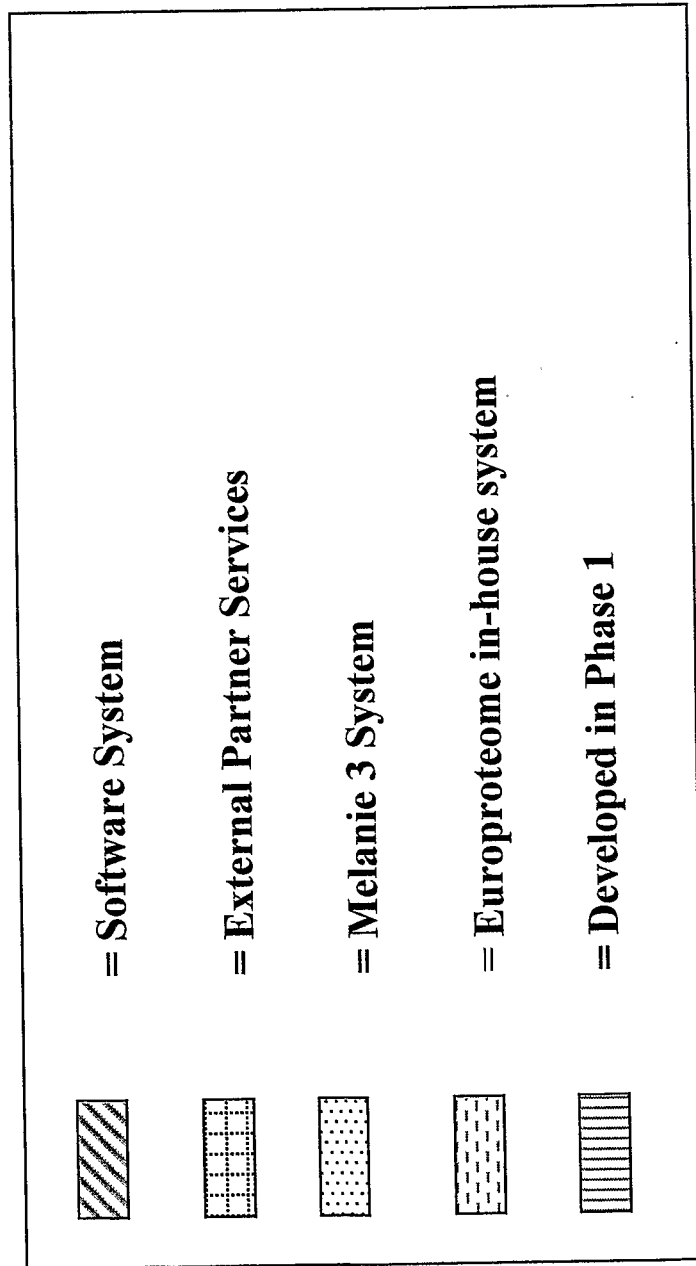


Fig 2d

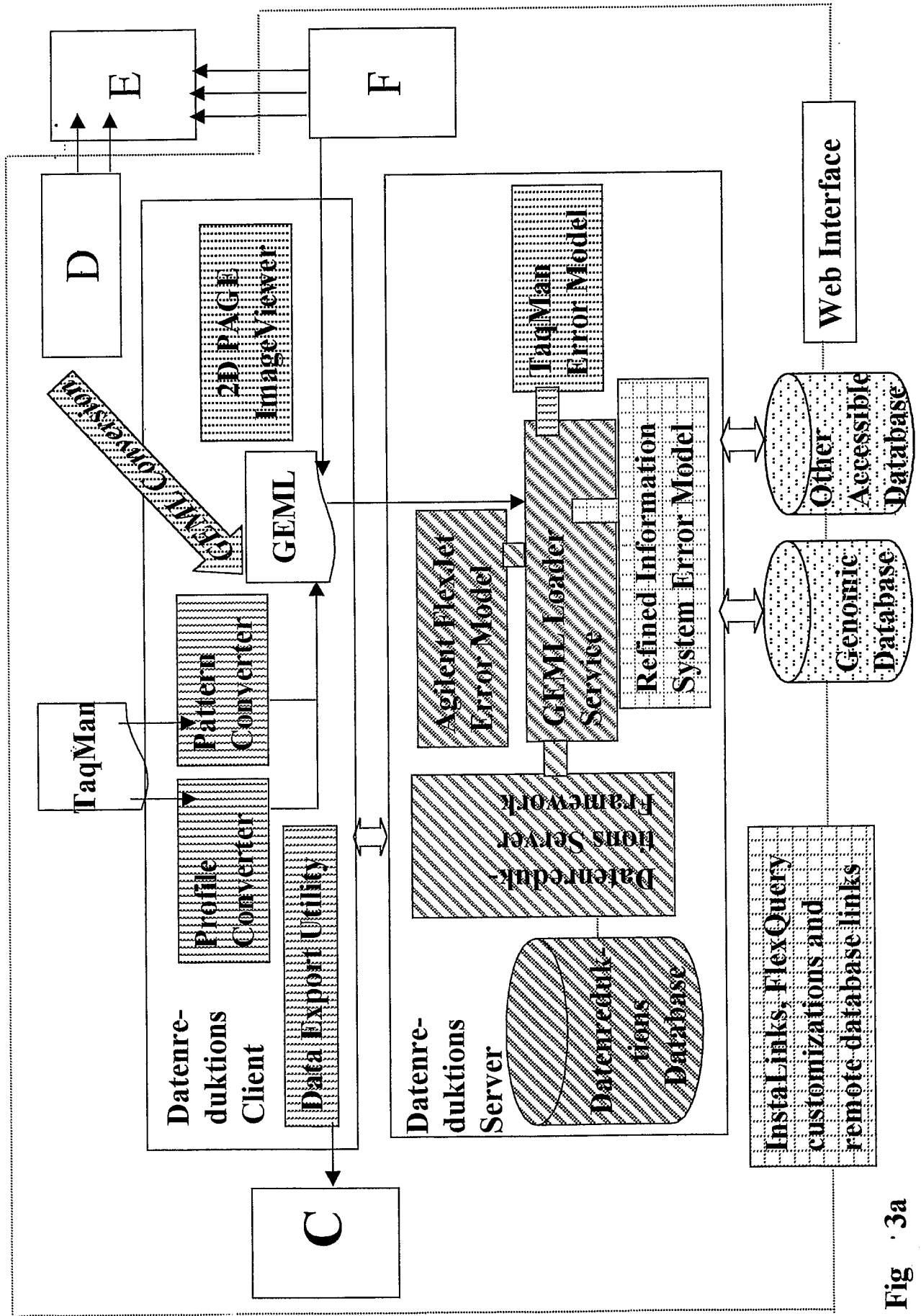


Fig . 3a

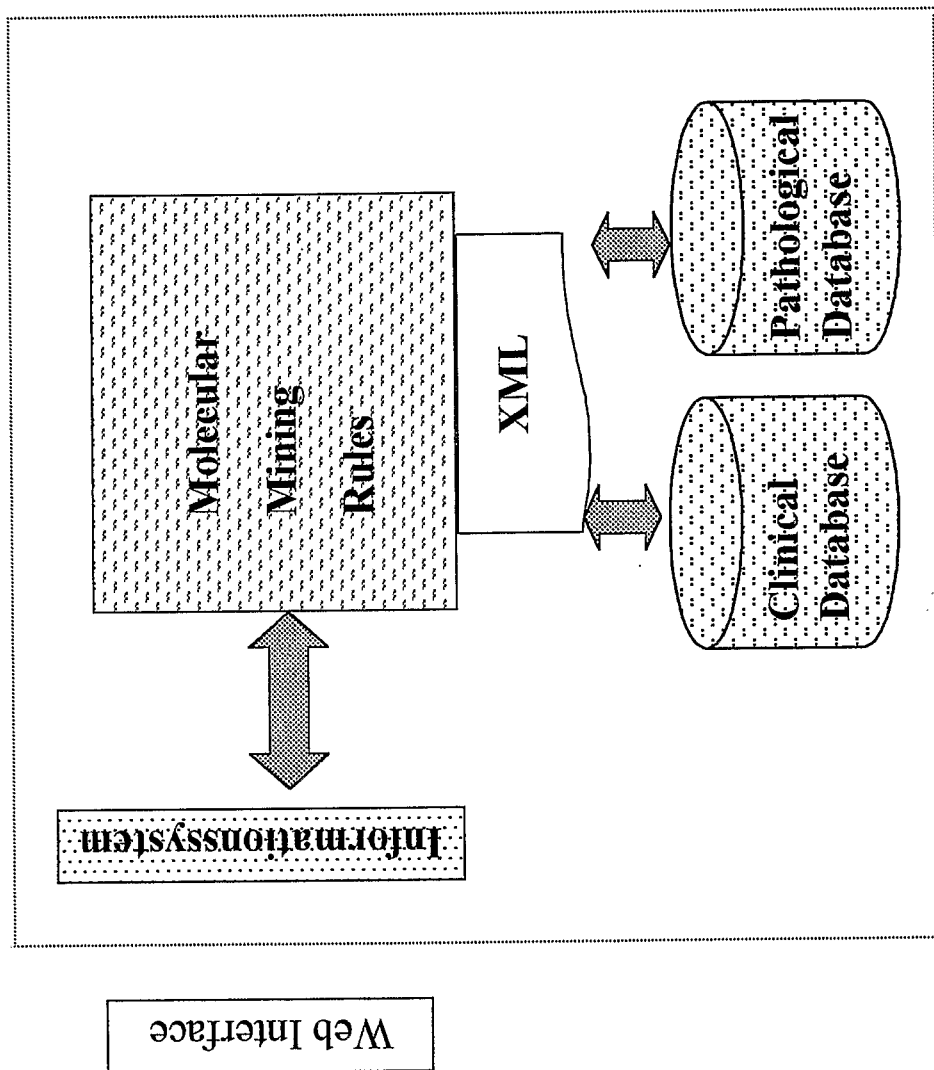


Fig. 3b

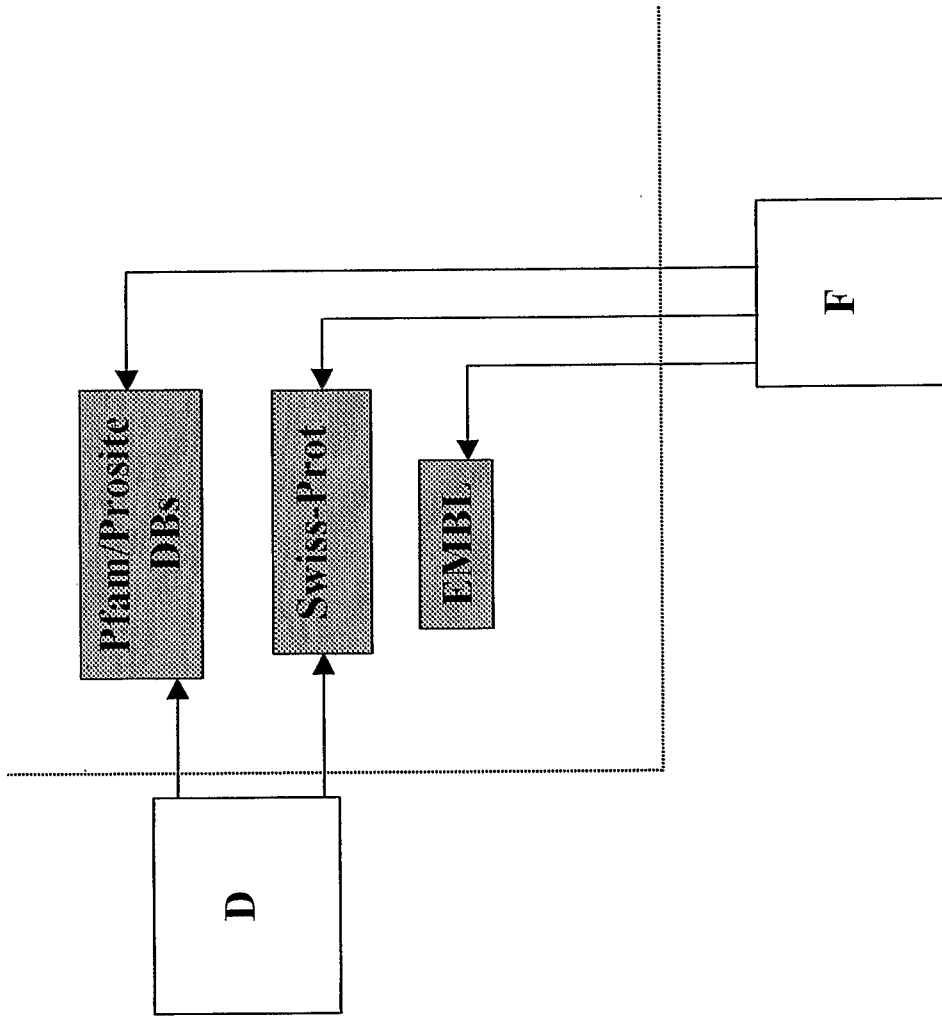


Fig 3d

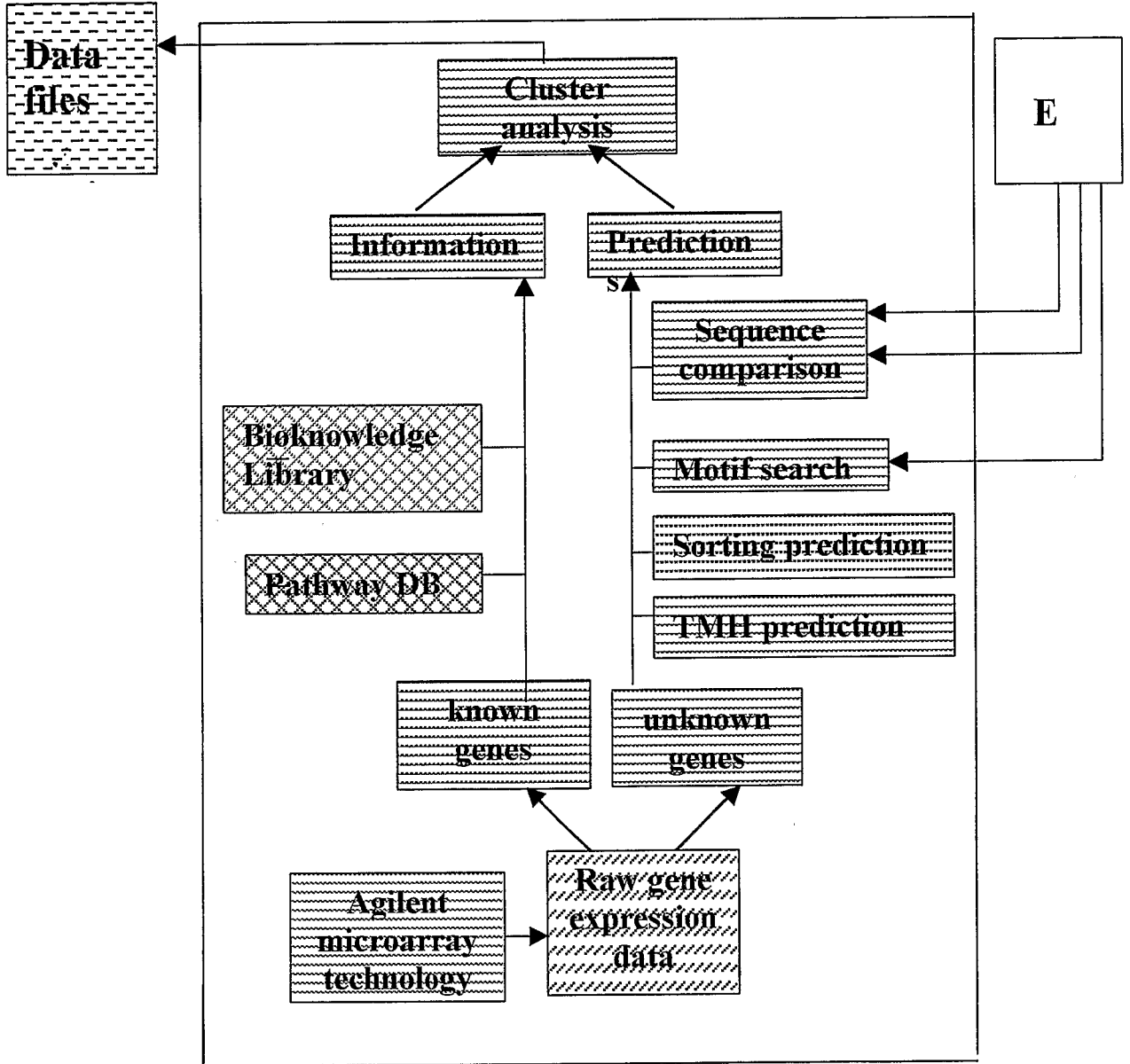


Fig 3e

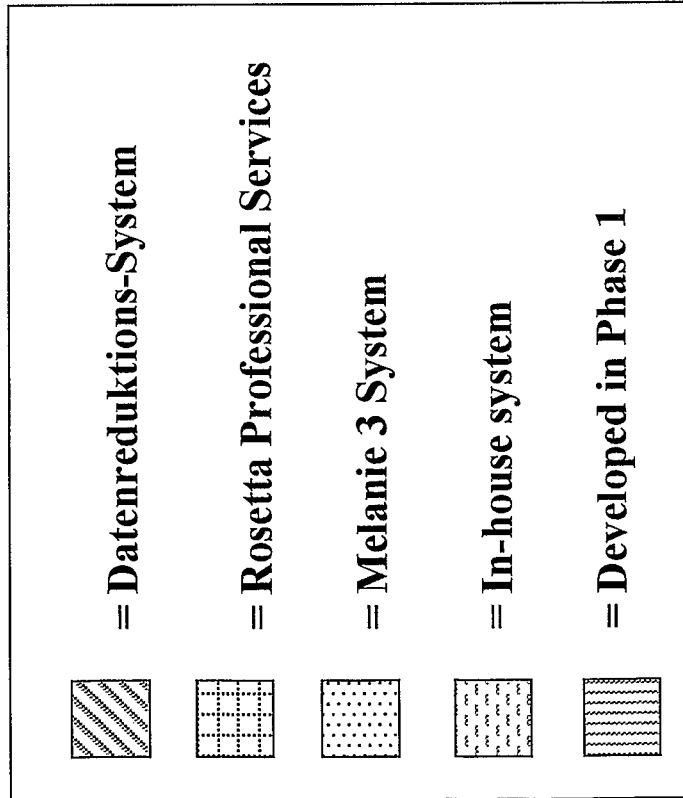
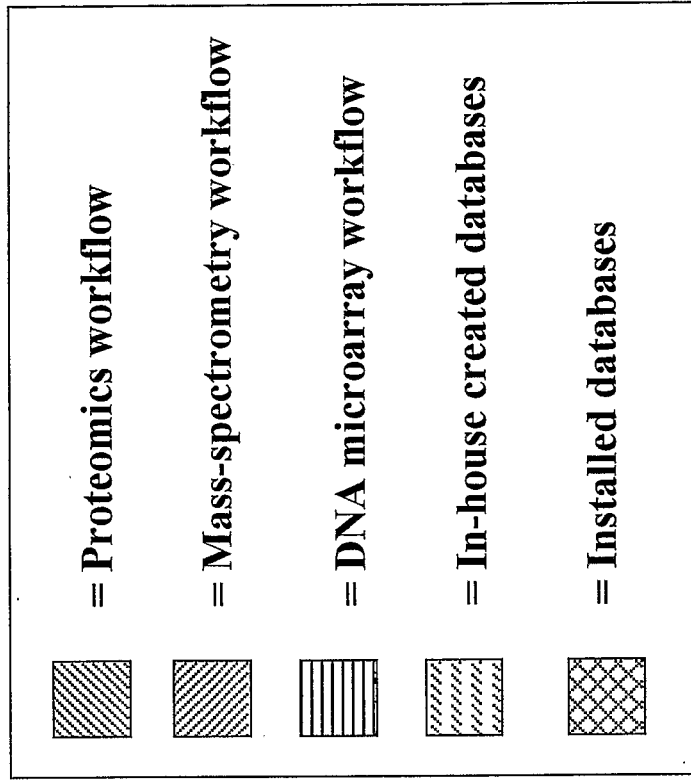


Fig 3f

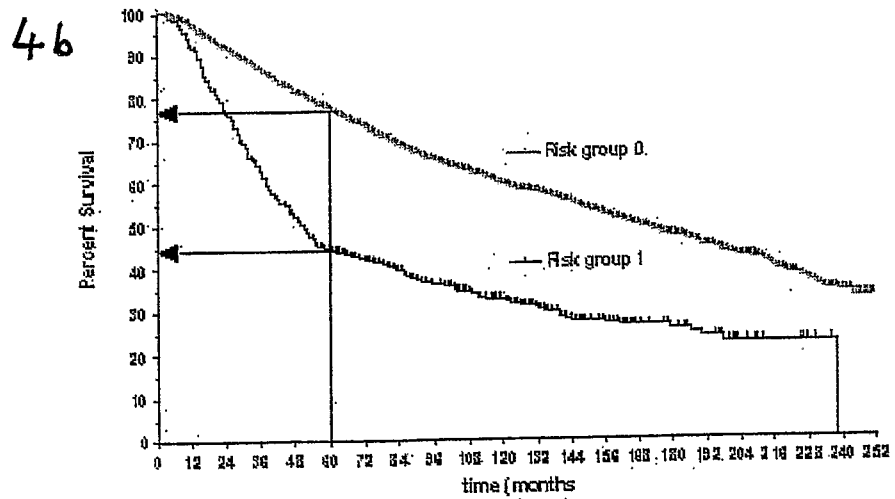
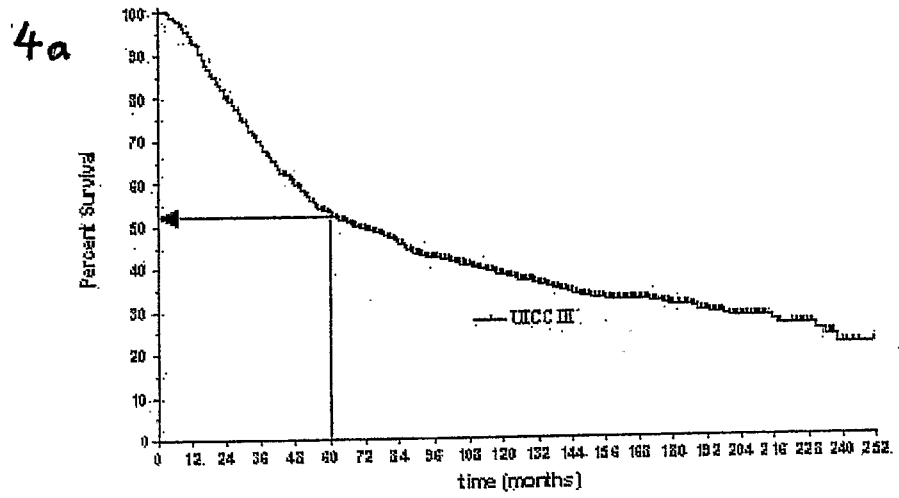


Fig. 4

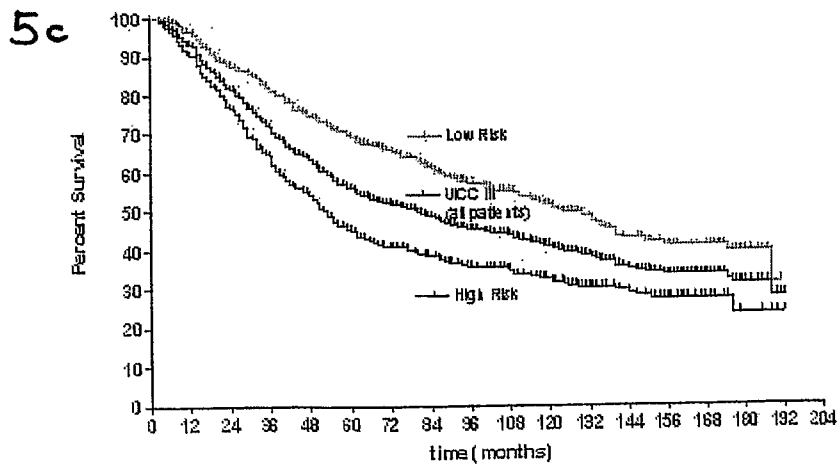
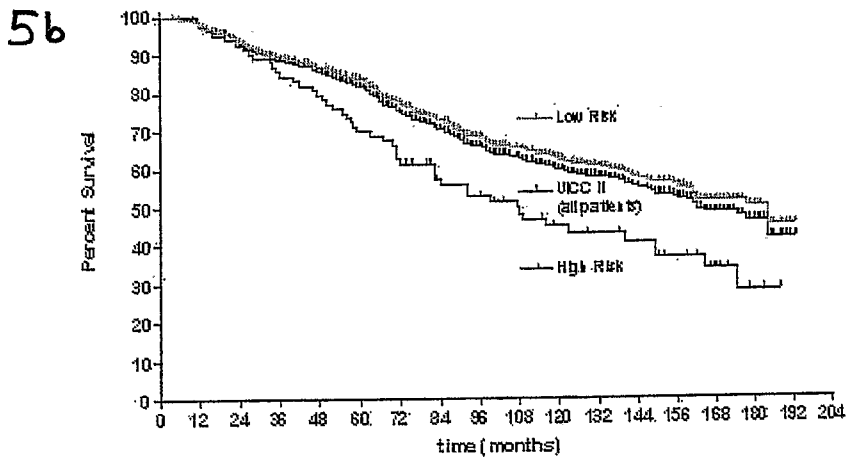
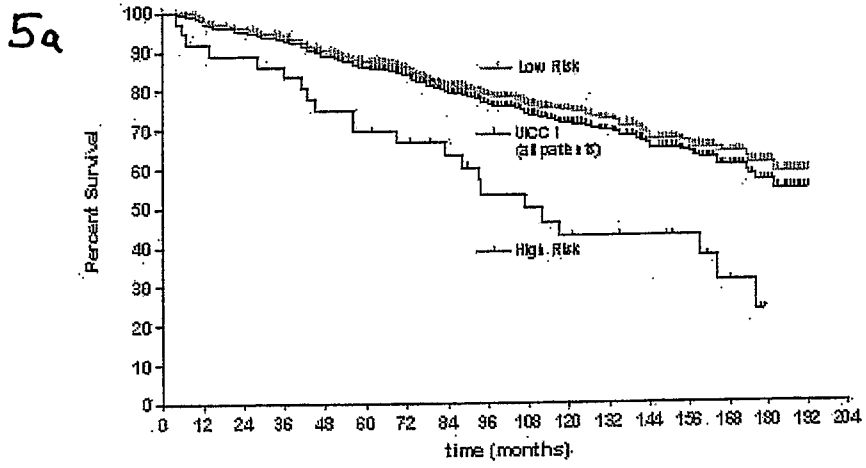


Fig 5

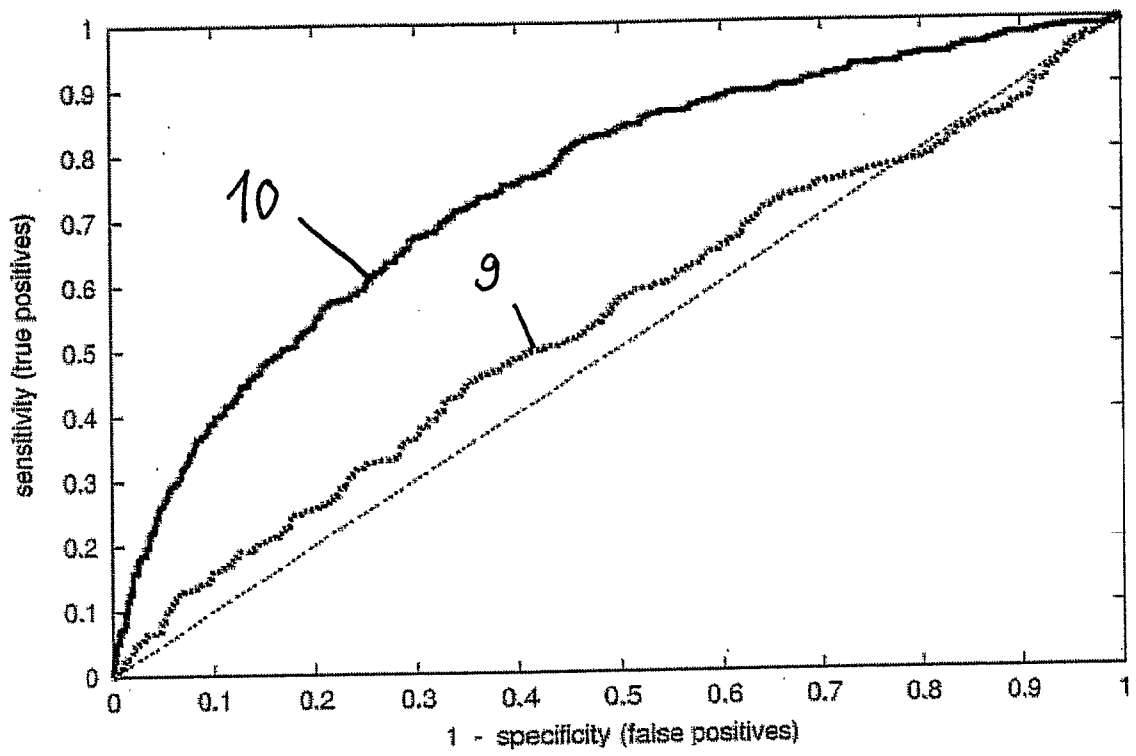


Fig 6

专利名称(译)	用于数据评估的方法和系统，相应的计算机程序产品和相应的计算机可读存储介质		
公开(公告)号	EP1423806A2	公开(公告)日	2004-06-02
申请号	EP2002797658	申请日	2002-08-30
[标]申请(专利权)人(译)	EUROPROTEOME		
申请(专利权)人(译)	EUROPROTEOME AG		
当前申请(专利权)人(译)	EUROPROTEOME AG		
[标]发明人	REYMOND MARC ANDRE		
发明人	REYMOND, MARC, ANDRE		
IPC分类号	A61B5/00 G06F17/30 G06F19/00 G06Q50/00		
CPC分类号	G16H50/20 G16H50/70 Y02A90/26		
优先权	10143712 2001-08-30 DE		
外部链接	Espacenet		

摘要(译)

本发明涉及用于数据评估的方法和系统，相应的计算机程序产品，以及相应的计算机可读存储介质，其可以特别地用作基于因特网的患者特异性预后系统。在这种情况下，可以整合临床，病理和分子生物学数据，并且所述数据可以与特定患者的相关预后相结合。因此，该系统使肿瘤学家能够根据特定的信息模式决定个体治疗。通过确定显著和次要变量，导致待评估数据量的明显减少，数据评估的加速和预后的改善，改善了预后的质量。