

André Rosenthal

## Die Nutzung des Humangenoms in der Onkologie

Lieber Mitja, liebe Inge,  
sehr verehrte Anwesende,

gerne bin ich der Einladung gefolgt, heute anlässlich Deines 90. Geburtstags, lieber Mitja, zu sprechen. Ich möchte aber nicht unerwähnt lassen, dass ich, als ich die Einladung in den Händen hielt, für einige Zeit daran denken musste, dass diese Einladung eigentlich Sina gilt. Ich habe eine besondere Briefmarke auf die Einladung geklebt und sie weitergeschickt, dort wo sie eigentlich hingehört. Es ist sehr schade, dass Sina heute hier nicht sprechen kann. Ich bin mir sicher, sie hätte viel und Spannenderes zu erzählen gehabt, aus Eurer gemeinsamen Zeit in der Hessischen Straße und danach. Viele Jahre habt ihr gemeinsam geforscht und auch manche Wissenschaftspolitik betrieben. Nicht immer ist alles so aufgegangen, wie ihr es erhofft hattet. Ich hoffe, Du gestattest mir, dass ich meinen Vortrag nicht nur Dir, lieber Mitja, sondern auch Sina widme. Mögen Dir noch viele weitere gesunde und produktive Jahre, zusammen mit Deiner lieben Inge, vergönnt sein.

### 1. Das menschliche Genom – Geschichte und Politik

In den letzten zwei bis drei Jahren wurde viel über das Human Genom Projekt (HGP) in den Medien berichtet. Einige verglichen die Entschlüsselung des menschlichen Genoms mit der Landung von Menschen auf dem Mond, andere mit der Erfindung des Buchdrucks im Mittelalter. Im Kontext mit den genetischen Programmen anderer wichtiger Modellorganismen wie der Labormaus *M. musculus*, der Fruchtfliege *D. melanogaster* und des Rundwurms *C. elegans* ist die Kenntnis der Sequenz des menschlichen Genoms eine einzigartige wissenschaftliche Errungenschaft, für die es keine Parallele gibt. Die vielfältigen Informationen, die im genetischen Programm des Menschen verschlüsselt sind, werden in den nächsten Jahren und Jahrzehnten nach und nach analysiert und verstanden werden. Die mittel- und langfristi-

gen Auswirkungen auf die Wissenschaften, insbesondere auf die Biologie und Medizin, aber auch auf die menschliche Gesellschaft, können zum heutigen Zeitpunkt nur erahnt werden. Sehr wahrscheinlich wird sich unser Leben in der ferneren Zukunft als Folge dieser wissenschaftlichen Großleistung stärker verändern, als wir heute voraussehen können. Wie bei allen großen wissenschaftlichen Durchbrüchen kann es bei der späteren Anwendung dieses neuen Wissens auch zu negativen Konsequenzen für den Einzelnen aber auch für bestimmte Gruppen in der Gesellschaft kommen. Dies hängt jedoch in erster Linie von den jeweiligen politischen Umständen ab und sollte durch kluges, vorausschauendes und verantwortliches Handeln der Gesellschaft und aller ihrer Mitglieder vermieden werden.

Die Sequenzierung des menschlichen Genoms war, technisch gesehen, extrem herausfordernd und beschäftigte tausende Menschen in sechs Ländern über fast eine Dekade. Der Wettbewerb zwischen dem öffentlichen akademischen HGP Consortium und Celera Genomics beschleunigte das Projekt wesentlich. Im Dezember 1999 wurde die Sequenz des ersten menschlichen Chromosoms – Chromosom 22 – von einem Team unter der Leitung des Britischen Sanger Centres fertig gestellt (1). Im Mai 2000 folgte dann das Chromosom 21, welches von einem Deutsch-Japanischen Konsortium sequenziert wurde (2). Ein großer Teil von Chromosom 21 und von den Chromosomen 8 und X wurde unter meiner Leitung am Institut für Molekulare Biotechnologie in Jena sequenziert und analysiert. Im Dezember 2001 wurde dann das Chromosom 20 wiederum vom Britischen Sanger Centre beendet. Alle drei Chromosomen wurden in der angesehenen Zeitschrift „Nature“ publiziert, wobei Chromosom 21 bezüglich der Vollständigkeit und Qualität der Sequenz sehr wahrscheinlich als Goldstandard dienen wird. Im Chromosom 21 verblieben nur drei Lücken von ungefähr 100 Kilobasen oder etwa 0,3% des langen Arms, während bei Chromosom 22 noch etwa 1 Million Basen oder 3% der Sequenz fehlen. Am 26. Juni 2000 wurde in internationalen Pressekonferenzen rund um den Erdball die Fertigstellung der Rohfassung des Humangenoms angekündigt, die dann am 15. Februar 2001 wiederum in „Nature“ publiziert wurde (3).

### **1.1 Der Wettlauf zwischen dem HGP-Konsortium und Celera**

Zwanzig Zentren nahmen an der Sequenzierung des menschlichen Genoms teil (Tab. 1). Sie entwickelten eine hierarchische Kartierungs- und Sequenzierungs-Strategie, die der Komplexität des menschlichen Genoms angepasst wurde. Diese Strategie umfasst mehrere Schlüsselmethoden (Abb. 1). Zunächst wurde die genomische DNA in mehrere hunderttausend kleinere

Fragmente, s.g. BACs („bacterial artificial clone“) und PACs kloniert. Jeder BAC/PAC-Klon enthält ca. 100–200 Kilobasen menschlicher DNA. Verschiedene humane BAC-Bibliotheken (Caltech B, C, D1, RPCI-1, -3, -4, -5 und RPCI-11) mit mehr als 1,4 Millionen individuellen Klonen wurden über die Jahre angelegt. Die DNA stammte von 24 anonymen menschlichen Spendern, wobei verschiedene ethnische Gruppen berücksichtigt wurden. Anschließend wurden 354510 individuelle BAC-Klone, die hauptsächlich von der RPCI-11 und der Caltech D1 Bibliothek stammten, mit Hilfe einer Fingerprinttechnik physikalisch kartiert. Dabei wurde für jeden der ca. 350.000 Klone ein rechter und linker Nachbar identifiziert und alle Klone den 23 Chromosomen genau zugeordnet (4). Dazu verdaut man die DNA jedes BAC-Klons mit dem Restriktionsenzym HindIII und misst die Größe der sich ergebenden Restriktionsfragmente durch Agarose oder Polyacrylamid-Gelelektrophorese. Die Restriktionsmuster sind physikalische Fingerabdrücke für jeden Klon. Der Vergleich aller Fingerabdrücke im Computer erlaubt dann, zehntausende verschiedene BAC-Klone miteinander zu vergleichen. Klone, die identisch sind, haben identische Fingerabdrücke. Klone, die überlappen, haben partiell identische Abdrücke. Zusätzlich wurden die Enden von mehr als 800.000 BAC-Klonen ansequenziert, wobei von mehr als 480.000 BAC-Klonen Sequenzen von beiden Enden (gepaarte Sequenzreads) erzeugt wurden. Diese relativ kurzen Sequenzen erleichterten wesentlich die richtige Zuordnung der BACs. Zum Schluss wurden die BAC-Contigs mit Hilfe von mehreren tausend STS-Marker auf die 23 menschlichen Chromosomen zurück kartiert. Dabei wurde eine Vielzahl von Methoden verwendet, z.B. kann man individuelle BAC's mit Hilfe der fluoreszenten *in-situ* Hybridisation (FISH) einzelnen Chromosomenbanden zuordnen. Anschließend wurde für jedes menschliche Chromosom ein Satz von BAC/PAC-Klonen ausgewählt, der bei minimaler Überlappung der Einzelklone die gesamte Länge des Chromosoms abdeckt. Man bezeichnet diesen Satz von Einzelklonen als „minimal tiling path“ eines Chromosoms. Der „minimal tiling path“ des gesamten menschlichen Genoms umfasst ca. 30.000 einzelne Klone. In der Sequenzierungsphase wurden dann für jedes einzelne Chromosom alle BAC/PAC-Klone des „minimal tiling path“ des entsprechenden Chromosoms vollständig mit Hilfe der Schrotschussmethode sequenziert. Für das gesamte Genom wurden insgesamt 29.298 individuelle, gut kartierte BAC-Klone (25.241 Klone von RPCI-11 und Caltech D1 und 4.057 Klone anderer Bibliotheken) sequenziert. Über 80% der Rohfassung wurde zwischen Mai 1999 und Juni 2000 generiert (Abb. 2). Zusätzlich wurden noch 5,7 Millionen Sequenzen

von menschlichen Plasmid-Bibliotheken erzeugt, die durch Klonierung der genomischen DNA von 24 anonymen Spendern entstanden. Jeder dieser Plasmide trägt ein humanes DNA-Insert von ca. 1.000 bis 2.000 Basen Länge. Mit Hilfe dieser Sequenzen können Sequenzvariationen und –polymorphismen zwischen den Individuen und ethnischen Gruppen im Genom identifiziert werden. Um eine hohe Genauigkeit zu erzielen, wurden insgesamt 23 Milliarden oder  $23 \times 10^9$  Basenpaare sequenziert, d.h. jede Base wurde im Durchschnitt 7,5fach unabhängig voneinander bestimmt.

Zum Zeitpunkt der Publikation der Rohfassung im Februar 2001 waren erst 897 Megabasen d.h. 25% des Genoms in hoher Qualität erfasst (3). Zum Zeitpunkt dieses Vortrags sind mehr als 80% des Genoms lückenlos und liegen in hoher Genauigkeit vor (Genauigkeit: 99,99% d.h. weniger als ein Fehler in 10.000 Basen).

Im Gegensatz zum akademischen Consortium verfolgte Celera Genomics eine andere Strategie, die man auch als gesamte Schrotschussmethode („whole shotgun sequencing strategy“) bezeichnet. Mit Hilfe von verschiedenen Restriktionsenzymen wird die genomische DNA zunächst in drei parallelen Ansätzen verdaut, wobei man darauf achtet, dass im Durchschnitt Fragmente von 2 Kilobasen, 10 Kilobasen und 50 Kilobasen Länge entstehen. Die Fragmentgemische werden dann in präparativen Agarosegelen aufgetrennt, und anschließend isoliert man die genomische DNA der gewünschten Länge aus der Agarose und kloniert diese in geeignete Plasmidvektoren. Anschließend wurden insgesamt 13,5 Millionen Klone aus der 2 Kb-Bibliothek, 10,8 Millionen aus der 10 Kb-Bibliothek und 2,8 Millionen Sequenzen aus der 50 Kb-Bibliothek sequenziert. Die 26 Millionen einzelnen Sequenzen sollten dann mit Hilfe einer eigens von Celera entwickelten Software im Computer wieder zusammengesetzt werden. Entsprechend dieser Strategie sollten die Sequenzpaare der größeren 10kb und 50kb-Klone eine eindeutige Assemblierung der 13,5 Millionen kurzen Sequenzen erlauben. Celera publizierte seine Version des menschlichen Genoms am 16. Februar 2001 in der Zeitschrift „Science“ (5).

Die Analyse der publizierten Daten hat nun ergeben, dass die Celera-Version des menschlichen Genoms mehr als 60% der Daten des akademischen HGP-Konsortiums enthält. Da das HGP-Konsortium im Gegensatz zu Celera seine Daten täglich im Internet publizierte, konnten die Wissenschaftler von Celera diese Daten herunterladen und bei ihrer Assemblierung verwenden. Weiterhin wurde nachgewiesen, dass Celera ohne die Daten des akademischen HGP-Konsortiums nicht in der Lage gewesen wäre, das Genom fehlerfrei zu assemblieren. Die Schrotschussmethode von Celera kann zwar bei kleineren Organismen wie z.B. Bakterien, nicht aber beim menschlichen Ge-

nom angewendet werden. Die Genome von Säugetieren enthalten zu über 50% mittel- und hochrepetitive Sequenzen, die im Computer in einem Schrottschuss-Ansatz nicht fehlerfrei zusammengesetzt werden können. Essentieller Bestandteil der Strategie des HGP-Consortiums ist zunächst die Reduktion der Komplexität des menschlichen Genoms um den Faktor  $30 \times 10^4$ . Das wird durch die Aufteilung des Genoms in ca. 30.000 kleinere Fragmente (BAC/PAC-Klone) erreicht. Die Sequenzierung jedes der 30.000 Fragmente erfolgte individuell und unabhängig von den anderen Fragmenten. Durch diese Strategie wurde der enorm störende Einfluss der mittel- und hochrepetitiven Sequenzen im Genom erheblich reduziert.

Im Verlauf des Jahres 2003 sollen alle menschlichen Chromosomen publiziert werden. Dazu müssen die noch vorhandenen Kartierungs- und Sequenzierungslücken geschlossen werden. Nimmt man Chromosom 21 als Goldstandard, so sollte die finale Sequenz des menschlichen Genoms nicht mehr als 300 Lücken aufweisen – ein Ziel, das schwer zu erreichen sein wird. Was am Ende des Humangenomprojektes zählt, ist nicht der Wettlauf zwischen dem HGP-Konsortium und Celera, sondern die Sequenz des menschlichen Genoms in der höchst möglichen Qualität. Es ist sehr wichtig zu erreichen, dass die Sequenz des menschlichen Genoms zusammen mit den Sequenz-Annotationen (Interpretationen) für jedes Individuum, für jede öffentliche oder private Institution und Firma ohne Restriktionen jedweder Art zugänglich ist und über viele Jahre und Jahrzehnte zugänglich bleiben wird. Es ist wichtig, dass die Annotation, d.h. die Interpretation zusammen mit der Sequenz dem Benutzer dargeboten wird. Erst die Kombination aus Sequenz und Interpretation, die sich im Laufe der Zeit ständig verändern wird, macht das Genom zu einem einzigartigen wissenschaftlichen Schatz.

## 1.2 Wieviele Gene sind im Genom codiert?

Längere Zeit hatten die Biologen angenommen, dass das genetische Programm des Menschen von mehr als 100.000 Genen gesteuert wird. Zwischen 1996-98 wurde durch Sequenzierung gezeigt, dass die Bäckerhefe *S. cerevisiae* ca. 6.000 Gene, der Rundwurm *C. elegans* ca. 19.000 Gene und die Taufliege *D. melanogaster* ca. 16.000 Gene aufweisen. Aufgrund der geringeren Komplexität dieser Organismen gegenüber höheren Säugetieren schlussfolgerte man, dass im menschlichen Genom ein Vielfaches an Genen enthalten sein müsste. Auch die Sequenzierung von kurzen cDNA-Fragmenten s.g. EST-Fragmenten in den 90er Jahren erhärtete scheinbar diese These. Aus vielen menschlichen Geweben hatte man dutzende cDNA-Bibliotheken hergestellt und Millionen EST-Sequenzen generiert. Vor allem die US-Firmen

Incyte Genomics und Human Genome Science hatten in diese Technologie investiert, um auf diese Weise möglichst eine große Zahl der exprimierten Gene oder mRNAs zu gewinnen. Nach Assemblierung dieser Fragmente erhielt man einen Satz von ca. 100.000 bis 200.000 Contigs, der sich nicht weiter reduzieren ließ. Ende der 90er Jahre stützten alle experimentellen Daten scheinbar die Annahme, dass im Kern einer menschlichen Zelle mehr als 100.000 Gene repräsentiert sind. Die Sequenzierung und Analyse der Chromosomen 21 und 22 zeigte dann zu Beginn des 21. Jahrhunderts, dass beide Chromosomen zusammen nur etwa 750 Gene aufwiesen, sehr viel weniger als ursprünglich angenommen. Da beide Chromosomen etwa 2% des gesamten menschlichen Genoms ausmachen, konnte man die Gesamtzahl der menschlichen Gene mit ca. 37.500 abschätzen (2). Diese Hypothese wurde von mir am 13. Mai 2000 auf dem Cold Spring Harbor Symposium „Genome Sequencing & Biology“ zum ersten Mal vertreten und löste eine wissenschaftliche Sensation aus. Es zeigte sich dann schnell, dass diese Hypothese auch durch Resultate anderer Gruppen gestützt wurde. Durch vergleichende Analyse des Genoms des Kugelfisches mit dem der Chromosomen 22 und 21 konnte die Gruppe um Jean Weissenbach in Paris eine Gesamtzahl von ca. 30.000 menschlichen Genen ableiten. Eine kritische Analyse der menschlichen ESTs-Daten von Phil Green's Gruppe in Seattle kam ebenfalls auf ca. 30.000-40.000 Gene. Auch die Analyse der Rohfassung des menschlichen Genoms zeigt, dass im menschlichen Genom weniger als 40.000 Gene enthalten sind (3).

Die Frage, wieviele menschliche Gene durch das Genom genau codiert werden, ist nicht leicht zu beantworten. Sehr wahrscheinlich wird es noch viele Jahre dauern, bis wir die Gesamtzahl aller Gene mit großer Genauigkeit bestimmt haben. Zunächst müssen die Sequenzen aller menschlichen Chromosomen möglichst lückenlos vorliegen. Dies wird im Laufe des Jahres 2003 der Fall sein. Gegenwärtig haben wir umfangreiche experimentelle Beweise für ca. 12.000 menschliche Gene. Für diese Gruppe liegen auch komplette menschliche mRNA's vor, so dass wir in der Lage sind, die Exon/Intron-Struktur und die Größe dieser Gene genau zu bestimmen. Weitere ca. 10.000 Gene können mit Hilfe von bioinformatischen Methoden aus der Genomsequenz vorhergesagt werden. Es gibt eine Vielzahl von bioinformatischen Methoden, die einzeln oder in Kombinationen angewendet werden, um Genvorhersagen zu treffen. Die Güte der Vorhersagen kann sehr verschieden sein, je nachdem, welche Methoden bzw. Methodenkombinationen verwendet werden. Oft werden *ab-initio* Methoden, die auf statistischen Modellen beruhen, mit anderen *in-silico* Verfahren kombiniert, bei denen das codieren-

de Potential neuer DNA-Sequenzen mit allen bereits bekannten Proteinen und DNA-Sequenzen verglichen wird. Dabei fließen auch Informationen über regulative Sequenzen (Spleißsignale, Promotersequenzen, CpG-Inseln) ein. Die Ähnlichkeit einer unbekanntes DNA-Sequenz mit einem bereits experimentell beschriebenen Protein- oder RNA-Abschnitt, dessen Sequenz in den internationalen Datenbanken erfasst ist, erlaubt es dann, ein hypothetisches Genmodell aufzustellen. Diese *in-silico* Genmodelle erstrecken sich oft nur über einen kleinen Teilbereich des hypothetischen Gens, nämlich über die Regionen, wo es Ähnlichkeiten zwischen den konservierten Proteindomänen gibt. Die *in-silico* Genmodelle sind zunächst reine Arbeitshypothesen und müssen durch aufwendige experimentelle Methoden (Klonierung, RT-PCR, *in-situ* Hybridisierung) bewiesen werden. Das heißt, während wir von ca. 22.000 Genen eine recht gute Vorstellung haben, liegen uns von weiteren 10.000-15.000 Genen nur sehr unvollständige und fehlerhafte Genmodelle vor. Menschliche Gene, die keinerlei Ähnlichkeit zu bekannten Genen und Proteinen aufweisen, können durch gängige bioinformatische Methoden nicht vorhergesagt werden. Nur *ab-initio* Methoden erlauben, Modelle solcher hypothetischen Gene vorherzusagen. Diese Methoden weisen jedoch eine hohe Rate von falsch-positiven Vorhersagen auf.

Die Bestimmung der genauen Zahl aller menschlichen Gene ist auch deshalb so schwierig, weil relativ viele Gene nur auf sehr niedrigem Niveau, zu einem ganz bestimmten Zeitpunkt und nur in bestimmten Zellen bzw. Geweben exprimiert werden. Diese Gene wird man experimentell nur dann finden, wenn cDNA-Bibliotheken dieser Gewebe bzw. Zellen hergestellt und sequenziert bzw. wenn Genmodelle mit Hilfe der RT-PCR in diesen Geweben überprüft werden. Dies ist ein sehr aufwendiges und langwieriges Unterfangen.

### 1.3 Vergleichende Analyse mit anderen Säugergenomen

Kürzlich wurde in „Nature“ die Rohfassung des Mausgenoms publiziert (6). Bald werden auch Rohfassungen der Genome der Ratte und des Schimpansen zugänglich sein. Die vergleichende Analyse der Genome verschiedener Säuger hat eine Reihe von Vorteilen. Nur auf diesem Weg wird man die Zahl und Struktur der menschlichen Gene relativ genau bestimmen können. Die Analyse des menschlichen Genoms im Vergleich mit Genomen von Wirbeltieren und wirbellosen Organismen reicht hierfür nicht aus. Zweitens wird der Vergleich des menschlichen Genoms mit dem von Maus und Ratte erlauben, die Vielfalt der regulativen Sequenzen in Säugergenomen zu entdecken und deren Funktionen aufzuklären. Dadurch wird es auch möglich sein, die prinzi-

piellen Fragen der Regulation der Genexpression besser zu adressieren. Und schließlich wird der Vergleich verschiedener Säugergenome insbesondere der Vergleich von Mensch und Primaten die Rolle von Sequenzvariationen bzw. Polymorphismen in Bezug auf ihre Funktionen aufhellen.

#### **1.4 Von der Sequenz eines Gens zu seiner Funktion**

Das HGP Projekt hat sehr viele mittel- und langfristige Implikationen, die weit über das Kernprojekt d.h. die eigentliche Sequenzierung der ca. 3 Milliarden Basen hinausgehen, die das genetische Programm einer menschlichen Zelle ausmachen. Die schwierige Frage war und ist, wie bestimmen wir, ausgehend von der Sequenz, die Funktion aller menschlichen Gene in der gesunden Zelle, aber auch im kranken Organismus. Dabei kann man zwei Wege beschreiten. Entweder versucht man, die Funktion eines Gens im gesunden Gewebe aus seiner Sequenz abzuleiten und tastet sich dann zur veränderten Funktion des mutierten Gens im Rahmen einer erblich bedingten oder mitbedingten Erkrankung vor. Oder man identifiziert zunächst Mutationen in menschlichen Genen, die die genetische Ursache bestimmter monogener Erkrankungen bzw. genetische Teilkomponenten multifaktorieller Erkrankungen darstellen. Nachdem man die Funktion eines Gens über seine verursachende Rolle bei einer Erkrankung eingegrenzt hat, versucht man dann, die Funktion des Gens in der gesunden Zelle aufzuklären. Für die meisten der ca. 30.000-40.000 verschiedenen menschlichen Gene und den daraus ableitbaren Transkripten sind genaue zelluläre Funktionen bisher nicht beschrieben worden. Für etwa eintausend Gene wissen wir, dass sie in mutierter Form monogene Erkrankungen auslösen. Aber auch bei diesen Genen kennen wir meistens nicht die genaue Funktion in der normalen Zelle. Die im Internet verfügbare Sequenz des menschlichen Genoms hat die Identifizierung von Krankheitsgenen für monogene Erkrankungen enorm vereinfacht. Vor zehn, fünfzehn Jahren dauerte die Suche nach einem Krankheitsgen viele Jahre und verschlang enorme Ressourcen. Heute ist das eine Sache von Wochen oder Monaten. Für die meisten der etwa tausend bekannten Krankheitsgene sind dutzende, manchmal auch hunderte Mutationen beschrieben worden. Humangenetiker versuchen, die verschiedenen Mutationen in einem bestimmten Gen mit den beobachteten Phänotypen der Erkrankung zu korrelieren.



### **1.5 Die Aufklärung der genetischen Ursachen multifaktorieller Erkrankungen erfordert neue Strategien**

Sehr viel schwieriger ist es, die Funktion derjenigen menschlichen Gene, Transkripte bzw. Genprodukte zu bestimmen, die nicht mit monogenen Erkrankungen assoziiert sind. Bisher ist es den Wissenschaftlern nicht gelungen, die genetischen Komponenten für die große Gruppe der multifaktoriellen Erkrankungen zu identifizieren. Zu dieser Gruppe zählen die meisten Volkskrankheiten wie z.B. Krebs, Diabetes, Alzheimer, Parkinson, Multiple Sklerose und die meisten Herz-Kreislauf-Erkrankungen. Sie werden durch eine Vielzahl von Faktoren verursacht. Neben vielen verschiedenen genetischen Faktoren (z.B. Mutationen oder Polymorphismen in Genen) spielen vor allem auch bisher nicht verstandene Umweltfaktoren eine wichtige Rolle. Vielversprechend sind experimentelle Strategien, bei denen in Modellorganismen wie z.B. der Maus und der Ratte ein bestimmtes Gen gezielt ausgeschaltet („knock-out“) wird. Anschließend untersucht man den Effekt des fehlenden Gens auf den resultierenden Phänotyp. Man kann auch mehrere Gene gleichzeitig ausschalten, bzw. Gruppen von Genen nacheinander einzeln „ausknocken“. Seit kurzem benutzt man kurze RNA-Moleküle, s.g. „antisense RNA“ oder „siRNA“ zum Ausknocken von Genen. Vor allem die siRNA-Technologie ist sehr vielversprechend. Bei niedrigen Eukaryonten wie der Bäckerhefe, des Rundwurmes und der Taufliede gibt es Forschungsprogramme, bei denen man systematisch alle Gene dieser Organismen einzeln ausschaltet. Ähnliche Programme werden gegenwärtig für die Maus verwirklicht. Man kann aber auch Gene in einem Modellorganismus überexprimieren. Auf diese Weise hergestellte transgene Tiere (Zebrafisch, Maus, Ratte) eignen sich ebenfalls sehr gut, den Effekt einzelner Gene zu studieren.

### **1.6 Vom Genom zum Proteom und zur Komplexität einer menschlichen Zelle**

Die Situation wird noch komplexer, wenn man bedenkt, dass durch gewebespezifisches Spleißen mehrere mRNA's oder Transkripte aus einem einzigen Gen entstehen können. Bei Säugetieren wie z.B. dem Menschen beobachtet man ca. 3 differentielle Spleißprodukte für jedes Gen. Bei einigen Genen wurden sogar mehrere dutzend verschiedene Spleissvarianten identifiziert. Das menschliche Transkriptom umfasst demnach etwa 100.000–200.000 verschiedene mRNAs, die an den Ribosomen in Proteine übersetzt werden. Im Anschluss an die Proteinsynthese (Translation) kann das Protein zusätzlich

noch durch verschiedene Zucker- oder Phosphatreste modifiziert werden. Diese posttranslationalen Modifikationen sind Signale für den zielgerichteten Transport von Proteinen oder Proteinvarianten an bestimmte zelluläre Orte. Der gesamte Raum aller Proteine (Proteom) umfasst demnach wahrscheinlich eine halbe Million verschiedener Eiweiße, von denen jedoch nur ein Bruchteil in jedem Zell- bzw. Gewebetyp exprimiert ist.

Zusammenfassend kann man sagen, dass durch die Aufklärung des menschlichen Genoms sehr deutlich geworden ist, dass starre und mechanistische Auffassungen ungeeignet sind, um die Komplexität von Lebensprozessen zu beschreiben. So ist die genetische und zelluläre Komplexität einer Zelle bzw. eines Organismus nicht allein durch die Anzahl der Gene bestimmt, die das Genom dieser Zelle bzw. des Organismus ausmachen. Der Mensch hat nur etwa doppelt so viele Gene wie der Rundwurm *C. elegans* und etwa dreimal so viele Gene wie die Taufliege *D. melanogaster*. Die Komplexität einer Säugerzelle muss durch andere Faktoren bestimmt sein. Sie wird auch nicht allein durch die Anzahl der Transkripte und Proteine bestimmt, die z.B. in einer menschlichen Zelle zu einem bestimmten Zeitpunkt exprimiert werden. Erst wenn man die enorme Zahl der möglichen Protein-DNA- und Protein-Protein-Wechselwirkungen berücksichtigt und die Art und Weise, wie Signale zwischen Zellen und von der äußeren Membran in den Kern geleitet werden, beginnt man, sich schrittweise an die Komplexität von Lebensprozessen anzunähern. Diese wird vor allem durch dynamische Prozesse in Raum und Zeit wie z.B. Gen- und Proteinexpression, Proteintransport, Realisierung der Proteinfunktion in bestimmten zellulären Kompartimenten und Signaltransduktion beschrieben. Komplizierte Wechselwirkungsmechanismen, -netzwerke und -kaskaden sind fast völlig unerforscht. Chip-basierte Technologien zum Profiling der Expressionszustände von RNA's und Proteinen bieten einen ersten Ansatz, sich dieser Komplexität zu nähern.

## 2. Sequenzvariationen und neue therapeutische Optionen

Prädispositionen zu Erkrankungen sowie beobachtete Nebenreaktionen auf bestimmte Medikamente werden hauptsächlich durch individuelle Unterschiede in der Sequenz unseres Genoms bzw. in der Sequenz bestimmter Gene verursacht. Durch die Sequenzierung des menschlichen Genoms wissen wir, dass etwa je 1.000 Basen eine einfache Sequenzvariation oder polymorphe Base vorkommt. In unserem Genom gibt es also etwa 3 Millionen solcher polymorpher Basen, die man auch als "Single Nucleotide Polymorphism" (SNP) bezeichnet. Da das HGP-Consortium und Celera Genomics DNA von

anonymen Spendern einsetzte, kann man aus dem Datensatz beider Teams derartige SNP's mit Hilfe geeigneter Softwareprogramme herausfiltern. Das akademische Konsortium poolte die DNA von 24 Spendern und identifizierte ca. 1,4 Millionen verschiedene SNP's (3,7), während die Daten Celera's von 6 Spendern herrührten und insgesamt 2 Millionen SNP's identifiziert wurden (5).

Nur ungefähr 150.000–300.000 dieser SNP's liegen in protein-codierenden Abschnitten menschlicher Gene oder in kritischen Promoterregionen und können deshalb zu Aminosäureaustauschen bzw. zu einer wesentlichen Veränderung der Expressionshöhe des entsprechenden Proteins führen. Mit Hilfe von hochdichten Genchips verschiedener Firmen wie z.B. Affymetrix oder Febit kann man die Verteilung von hunderten oder gar tausenden SNP's zwischen Individuen und bestimmten Patienten, bei denen Nebenwirkungen beobachtet wurden, aber auch zwischen ethnischen Gruppen und Populationen in der Keimbahn aber auch in bestimmten Geweben untersuchen. Wenn man die SNP-Muster dann mit den klinischen Daten z.B. über die Wirkung oder Nichtwirkung eines Medikaments oder über beobachtete Nebenwirkungen korreliert, kann man den Einfluss des Genotyps eines Individuums oder einer bestimmten Subpopulation auf z.B. wichtige therapeutische Medikamente bestimmen. Diese auch als Profiling von Patienten bezeichneten Strategien haben ein sehr großes Potential und ermöglichen in der Zukunft, Medikamente in einem Maße patientenspezifisch zu dosieren, wie es heute nicht möglich ist. Außerdem könnte man vor Beginn einer medikamentösen Therapie diejenigen Patienten identifizieren, die von dem Medikament therapeutisch nicht profitieren oder gar Nebenwirkungen zu erwarten hätten. Für diese Gruppe von Patienten könnten frühzeitig alternative medikamentöse Therapien berücksichtigt werden. Das Profiling von Patienten sollte auch bei den sehr aufwendigen klinischen Prüfungen von neuen Medikamenten in Phase II und III eingesetzt werden, da diese zielgerichteter, schneller und ressourcenschonender durchgeführt werden könnten. Die Stratifikation von Populationen eröffnet auch vielfältige Möglichkeiten beim Design neuer Medikamente und Therapien. Diese neuen Strategien in der Diagnostik, Prävention und Therapie werden in der mittel- und langfristigen Zukunft zu einem radikalen Wandel der Medizin führen (Abb. 3).

### **3. Krebs und Krebsgene**

#### **3.1 Prävalenz von Tumoren**

Krebs ist nach den Herz-Kreislauf-Erkrankungen die zweithäufigste Todesursache. In den westlichen Industrienationen erkranken jedes Jahr ca. 10 Mil-

lionen Menschen an Krebs. Die verschiedenen Formen von Krebs verursachen jährlich ca. 4-5 Millionen Todesfälle in der westlichen Welt. Tab. 2 zeigt die Häufigkeit und die 5-Jahresüberlebensraten für die wichtigsten soliden Tumore für die USA im Jahr 2001. Prostata-, Brust-, Lungen- und Colonkarzinom sind mit ca. 198.000 bis 170.000 Neuerkrankungen pro Jahr in den USA die häufigsten Tumoren. Während Colon-, Brust- und Prostatakrebs 5-Jahresüberlebensraten von ca. 75-90% aufweisen, zeigt das Lungenkarzinom nur eine geringe Überlebensrate von 14%. Das Pankreaskarzinom ist mit einer 5-Jahresüberlebensrate von 4% noch tödlicher. 96% aller neu diagnostizierten Patienten mit Pankreaskarzinom versterben in den ersten sechs Monaten nach Diagnose.

Für fast alle soliden Tumore gilt, dass existierende Therapien oft nur eine limitierte Effektivität aufweisen. Speziell für die sehr aggressiven Tumoren der Bauchspeicheldrüse, der Lunge, des Eierstocks, aber auch für bestimmte Gehirntumoren (Gliome) und für das Nierenkarzinoms gibt es keine wirksamen Therapien.

### 3.2 Gegenwärtiger Stand der Krebs-Therapie

Solide Tumore werden gewöhnlicherweise durch chirurgische Eingriffe entfernt. Im Anschluss wird oft eine Chemotherapie - oft auch in Kombination mit Radiotherapie - durchgeführt. Bei einigen Tumoren erfolgt die Radio- oder Chemotherapie auch vor dem chirurgischen Eingriff.

Die Entwicklung eines neuen Medikamentes zusammen mit der präklinischen und klinischen Entwicklung dauert in der Regel 15-20 Jahre und kostet im Durchschnitt 300-500 Millionen US\$ (Abb. 4). Die Medikamenten-Entwicklung stellt eine Pyramide dar. Die Erfolgsrate, gemessen an den ursprünglich initiierten Forschungsprojekten, beträgt ca. 3%.

Im Laufe der letzten drei Jahrzehnte ist eine Vielfalt von chemotherapeutischen Agenzien durch die Pharmazeutische Industrie entwickelt worden (Abb. 5). Darunter befinden sich klassische zytotoxische Medikamente wie alkylierende Agenzien, Anti-Tumor-Antibiotika, Antimetaboliten, Pflanzenalkaloide, chemopräventive Agenzien aber auch verschiedene Hormone (z.B. Antiöstrogene). In den letzten zwanzig Jahren sind auch neue Therapie-strategien wie Immuntherapien und monoklonale Antikörper entwickelt worden. Von besonderem Interesse sind gegenwärtig vor allem humane monoklonale Antikörper (mAB), die durch transgene Mäuse erzeugt werden. Derartige Technologien und Produkte werden durch Firmen wie Abgenix, Medarex, Genmab und andere entwickelt. Aber auch humanisierte Antikörper und chimäre Maus/Mensch-Antikörper haben ein großes Potential und

sind von Firmen wie Genentech und Centocor erfolgreich in den Markt gebracht worden. Von Firmen wie CAT, Morphosys und Dyax werden sehr intensiv auch „Phage Display“-Technologien vorangetrieben, mit denen man hochspezifische Binder, s.g. „single chain“ Antikörper (scAB), erzeugen kann. Mit Hilfe einer weiteren Technologie kann man auch bispezifische Antikörper herstellen. Diese docken auf der einen Seite hochspezifisch auf der Tumorzelle an, besitzen am anderen Ende aber noch eine zweite Bindungsstelle für z.B. CD3, über die dann körpereigene Killerzellen wie z.B. Lymphozyten zur Tumorzelle mit dem Ziel rekrutiert werden, diese zu töten. Eine Reihe von Firmen wie z.B. Mikromet entwickeln derartige Strategien. Bispezifische Antikörper befinden sich noch in der Entwicklung und sind nicht im Markt zugelassen.

Circa 14 verschiedene mAB sind gegenwärtig durch die FDA zugelassen. Darunter sind mit Herceptin, Campath, Rituxan und Mylotarg vier Antikörper, die zur Therapie von Leukämien, Lymphomen und einigen soliden Tumoren wie z.B. Her2neu-positiven Mammakarzinomen zugelassen sind (Tab. 3). Vier weitere Antikörper, die hochspezifisch an Tumorzellen binden, diese aber über die mitgebrachte radioaktive Strahlung töten sollen, sind noch in der klinischen Testung.

### **3.3 Tumorigenese, Genetische Analyse und Krebsgene**

Krebs kann sich in fast jedem Teil des Körpers, jedem Gewebe und Zelltyp entwickeln. Krebs ist durch unkontrolliertes Zellwachstum gekennzeichnet. Normalerweise reproduzieren sich Zellen durch Zellteilung mit einer Geschwindigkeit, die im Gleichgewicht mit dem natürlichen Zelltod (Apoptose) ist. Auf diese Weise werden abgestorbene Zellen ersetzt. Tumorzellen haben ihre Fähigkeit zur Apoptose verloren und teilen sich ununterbrochen.

Bei den verschiedenen Krebsformen unterscheidet man familiäre und sporadische Tumoren. Familiäre Tumoren sind relativ selten, machen nur 5% aller Krebserkrankungen aus und werden durch Mutationen in bestimmten Genen in der Keimbahn verursacht. Mutationen in den Genen BrCa1 und BrCA2 bewirken z.B. bestimmte Formen des Mammakarzinoms, die wie andere erblich bedingte Erkrankungen von einer Generation auf die nächste vererbt werden. Man erkennt erblichen Brustkrebs daran, dass in einer Familie mehrere Frauen in verschiedenen Generationen am Mammakarzinom erkrankt bzw. verstorben sind. Es gibt auch eine Reihe von Colonkarzinomen, die erblich sind und durch Läsionen in bestimmten Genen verursacht werden. Im Gegensatz dazu sind sporadische Tumoren mit ca. 95% viel häufiger. Sie werden durch sequentielle Mutationen in Körperzellen in einer Vielzahl von

Genen bewirkt. Zufälligerweise verändert eine dieser Mutationen die Funktion eines kritischen Gens und führt zur Entstehung eines Zellklons, der einen bestimmten Wachstumsvorteil gegenüber den Nachbarzellen aufweist. Das Resultat ist ein expandierender Zellklon (Abb. 6). Weitere Mutationen in anderen Genen führen zu Wellen klonaler Expansionen und zu Zellen, die neben Wachstumsvorteilen zusätzliche Fähigkeiten erwerben. So können diese Zellen durch Abbau der extrazellulären Matrix (ECM) in die umgebenden Gewebe eindringen, die Basalmembran durchbrechen (Abb. 6) und sich über das Blut- und Lymphsystem an anderen Stellen des Körpers absiedeln, indem sie wiederum Basalmembranen durchdringen und Tochtergeschwülste, s.g. Metastasen bilden. Dieser Prozess wird als Metastasierung bezeichnet und ist durch die Veränderung einer Vielzahl von zellulären Funktionen gekennzeichnet. Bedenkt man die Komplexität dieses Vorgangs, könnten sehr viele verschiedene Gene beteiligt sein.

Genetische Analysen solider Tumoren mit Hilfe von molekulargenetischen Methoden, die man als Verlust der Heterozygotität ("loss of heterozygosity" oder LOH) und Vergleichende Genomhybridisation ("comparative genome hybridisation" oder CGH) bezeichnet, haben gezeigt, dass viele Regionen des Genoms in Tumoren amplifiziert oder verloren gegangen sind. Hunderte von Genorten (Loci) sind an der Tumorentstehung, -progression und -metastasierung beteiligt. LOH- und CGH- Analysen haben leider den Nachteil, dass ihre Auflösung sehr gering ist. Die Eingrenzung der kritischen Genloci gelingt deshalb leider nur mit einer Auflösung von 5-50 cM. Manchmal sind ganze Chromosomenarme in Tumoren aberrant. Fast ein Drittel des menschlichen Genoms ist in Tumoren genetisch instabil und deshalb in die Karzinogenese involviert. Es hat sich als sehr schwierig erwiesen, Tumorsuppressoren und Onkogene mit Hilfe von Positionalen Klonierungsverfahren auf der Basis von LOH und CGH zu identifizieren, die mit der Entstehung und Progression von sporadischen Tumoren assoziiert sind. Oft sind in der kritischen Region mehrere hundert Kandidatengene. Selbst wenn die kritische Region auf wenige cM reduziert wurde, sind immer noch zu viele Gene auf ihre Assoziation mit Tumorentstehung und -progression zu testen. Man kann versuchen, die Auflösung von CGH zu erhöhen, indem man BAC oder YAC Klone aus den kritischen Chromosomenregionen systematisch auf Objektträger aufspottet, die man für die CGH-Analysen einsetzt. Diese Methode, auch als Matrix-CGH bezeichnet, ist zwar konzeptional elegant, erweist sich aber technisch als sehr schwierig. Eine andere Option ist die Verwendung von hochdichten cDNA-Arrays für die CGH. Aber auch die LOH-Analysen werden von der Verfügbarkeit der Genomsequenz profitieren, da viele neue allelspezifische LOH-Marker leicht identifiziert werden können, die dazu beitragen, die Auflösung in einer kritischen LOH-Region zu erhöhen.

Ein anderes Problem ist, dass viele tumor-assoziierte Gene keine detektierbaren molekularen Läsionen wie z.B. homozygote Deletionen beider Allele oder heterozygote Mutationen aufweisen, die den offenen Leserahmen (ORF) verändern und zu verkürzten nichtfunktionalen Proteinen führen. Ein anderer wichtiger Aspekt moderner Tumorgenetik ist, dass viele tumor-assoziierte Gene einen Verlust der Expression zeigen, der nicht durch Mutationen oder Deletionen bewirkt wird, sondern aufgrund von Hypermethylierungen der Promotoren zustande kommt. Aufgrund epigenetischer Mechanismen sind viele potentielle Krebsgene heute experimentell „unsichtbar“ und können nur über aufwendige systematische Methylierungsstudien der Kandidatengene „sichtbar“ gemacht werden.

Trotz dieser grundsätzlichen Probleme sind eine Reihe wichtiger Tumorsuppressor-Gene identifiziert worden, wie z.B. *p53*, *p16*, *RBI*, *APC*, *MLH1*, *MSH2*, *PTEN*, *WT1*, *ATM*, *BrCa1/2*, *PTCH*, *CDH1* und *CDKN1C*, um nur einige zu nennen. In mutierter Form verursachen einige dieser Gene familiär vererbte Tumoren. Mutationen in diesen Tumorsuppressoren sind aber auch in sporadischen soliden Tumoren beobachtet worden. Außerdem wurden mehr als 100 dominante Onkogene identifiziert, viele davon in hämatologischen Tumoren wie Lymphomen und Leukämien. Mit Hilfe der fluoreszenten *in-situ* Hybridisation (FISH) konnten die Bruchpunkte in diesen hämatologischen Tumoren relativ leicht identifiziert und kloniert werden.

Die meisten tumor-assoziierten Gene sind bisher jedoch noch nicht identifiziert worden. Man nimmt an, dass einige hundert Gene in die Tumorigenese involviert sind und potentielle molekulare Angriffspunkte (Targets) in der Onkologie für die Entwicklung neuer Therapeutika darstellen. Diese Schätzung basiert auf der morphologischen, histopathologischen und genetischen Heterogenität solider Tumoren sowie auf der Komplexität zellulärer Prozesse der Tumorprogression und Metastasierung. Nur ein paar Dutzend dieser Gentargets konnten in den letzten Jahren gefunden werden.

#### **4. Einsatz genomischer und genetischer Strategien führt zu neuen Krebs-Therapeutika**

Es gibt eine Reihe genomischer Strategien, um neue Targets in der Onkologie zu entdecken. Eine Möglichkeit ist, mit Hilfe der Sequenz des menschlichen Genoms Listen plausibler Kandidatengene zu generieren, indem man nach Paralogen bekannter Krebsgene sucht (8). Man kann aber auch nach Genen suchen, die z.B. in DNA-Reparatur, Zellzyklus und Apoptose involviert sind. Modellorganismen wie die Bäckerhefe und der Rundwurm *C. elegans* sind dafür sehr gut geeignet, da man derartige Gene in diesen Organismen sehr gut

studieren kann. Nachdem Gene in diesen Modellorganismen charakterisiert wurden, sucht man mit Hilfe geeigneter bioinformatischer Verfahren im menschlichen Genom nach den entsprechenden orthologen Genen. Eine weitere wertvolle Strategie ist die Suche nach Genen, die bei der Bildung neuer Blutgefäße beteiligt sind. Dieser Prozess - auch als Angiogenese bezeichnet - ist für die Bildung und die Aufrechterhaltung von Tumoren wichtig. Die Blockierung der Angiogenese (Antiangiogenese) stellt einen neuen Ansatzpunkt zum rationalen Design neuer Therapeutika dar.

Die Entwicklung neuer therapeutisch wirksamer Inhibitoren auf der Basis kleiner organischer Moleküle oder auf Antikörperbasis, die entweder die Apoptose in Tumorzellen anschalten oder verstärken oder die Angiogenese unterdrücken, ohne gleichzeitig toxisch zu sein, stellt eine große Herausforderung für die pharmazeutische und biotechnologische Industrie dar.

Eine andere vielversprechende Strategie ist es, Proteintargets zu identifizieren, die in größerer Menge auf der Oberfläche von Tumorzellen oder in der Tumorzelle exprimiert sind, nicht jedoch oder in sehr geringem Umfang auf oder in den normalen Epithelien bzw. Zellen. Die Spezifität der Targetexpression und der Grad der differentiellen Expression zwischen Tumor- und Normalzelle determinieren das therapeutische Fenster eines auf diese Weise identifizierten Targets. Proteintargets, die auf der äußeren Membran der Tumorzelle exprimiert sind, können mit humanen monoklonalen Antikörpern angegriffen werden. Man versucht Antikörper zu entwickeln, die hochspezifisch an den Rezeptor binden, aber gleichzeitig die Signalübertragung ins Innere der Zelle modulieren. Antikörper wie C225/Erbitux gegen EGFR oder Herceptin gegen ErbB2/Her2neu sind Vertreter dieser Klasse und sind gegen verschiedene Vertreter der Tyrosinkinase-Rezeptoren (TKR) gerichtet. Targets im Zytosol einer Zelle können mit Hilfe von Inhibitoren mit geringer Molekülgröße blockiert werden. Ein Beispiel dieser Klasse ist das neue Medikament „Iressa“ von AstraZeneca, das an die ATP-Bindungsstelle der Tyrosinkinase von EGFR bindet und dadurch die Phosphorylierung von Zielproteinen weiter stromabwärts verhindert. Abb. 7 zeigt neben Iressa eine Reihe weiterer neuer Medikamentenkandidaten auf der Basis kleiner Moleküle, die ebenfalls die Tyrosinkinaseaktivität einiger Mitglieder der Familie der Tyrosinkinase-Rezeptoren EGFR und Her1-4 inhibieren und sich gegenwärtig in der klinischen Testung befinden.

Neue Gentargets müssen definitionsgemäß zu bestimmten Proteinklassen gehören, für die in der Vergangenheit der Nachweis erbracht wurde, dass gegen sie wirksame Arzneimittel entwickelt werden konnten. Diese Targets bezeichnet man als „drugable“. Dazu gehören bestimmte Transmembranproteine wie die große Familie der G-gekoppelten Rezeptoren, die Familie der Ty-



rosinkinase-Rezeptoren, bestimmte sezernierte Proteine und ihre Rezeptoren, Kinasen (hier wiederum bevorzugt Tyrosinkinasen), Phosphatasen und bestimmte Proteasen wie z.B. Serinproteasen oder Matrix-Metalloproteasen (MMPs). Ein Gentarget sollte notwendigerweise in einer signifikanten Zahl der Tumorpatienten hochreguliert sein und ein therapeutisch nutzbares Fenster der Expression aufweisen. Der Wert eines Targets steigt jedoch erheblich mit dem Grad seiner Validierung. Es ist notwendig, in zellulären Assays zu zeigen, dass die Herunterregulation eines Targetproteins durch mAB, antisense RNA oder siRNA die Proliferation, Invasion oder Motilität der Zellen beeinflusst. Überexpression eines Targets sollte zu einer Reduktion der Apoptose im zellulären System führen, während Herunterregulation die Apoptose stimulieren sollte. Die Bindung eines Arzneimittelkandidaten (mAB, Inhibitoren mit kleinem Molekulargewicht, antisense/RNAsi/Ribozym) an das Target sollte in entsprechenden orthotopischen und subkutanen Mausmodellen signifikant das Tumorwachstum inhibieren.

## **5. Die Vision von metaGen**

metaGen ist eine biopharmazeutische Forschungs- & Entwicklungsfirma, die innovative genomische, genetische und molekular-pathologische Technologien für die Entwicklung neuer Verfahren für die Diagnose und Therapie von soliden Tumoren einsetzt. Das F&E Programm von metaGen ist auf das Mamma- und Prostatakarzinom sowie auf Tumore der Lunge und des Colon gerichtet. Außerdem werden noch Eierstock- und Blasen Tumore bearbeitet.

Das Verfahren von metaGen zur Identifikation und Selektion validierter Targets in der Onkologie umfasst eine Technologieplattform, bei der eine Reihe wichtiger Verfahren und Methoden kombiniert werden. Zunächst wird mit Hilfe eines klinischen Netzwerks, das über ganz Deutschland verteilt ist, in allen sechs Tumorindikationen frisches Tumormaterial und korrespondierendes Normalgewebe zur Verfügung gestellt. Unter Nutzung der Laser-Mikrodissektion, deren Grundprinzip in Abb. 8 dargestellt ist, werden aus den sehr heterogenen Tumoren die Tumorzellen und normale Epithelzellen vom gleichen Patienten gewonnen und angereichert (Abb. 9). Dazu wird in einer Serie von 10-20 Gefrierschnitten, die vom gleichen Patienten stammen und ca. 6 Mikrometer dünn sind, die Tumorfoci bzw. das normale Epithel mit Hilfe einer "Maus" im Mikroskop angezeichnet. Mit Hilfe eines UV-Lasers, der über das Objektiv eines Umkehrmikroskops auf den Schnitt eingekoppelt wird, werden dann alle markierten Tumorfoci bzw. das normale Epithel automatisch ausgeschnitten. Um ca. 50 Tumore und das korrespondierende Normalgewebe zu mikrodissektieren, braucht man mit 2 Laser-Mikrodissek-

tions-geräten ca. ein Jahr. Die Mikrodissektion von normalen Epithelien ist besonders schwierig und dauert in der Regel dreimal länger als die Mikrodissektion von Tumorfoci. Besonders aufwendig und schwierig ist die Mikrodissektion von Epithelien des Pankreas- und Lungenkarzinoms. Bei vielen Tumoren (Mamma-, Ovarial-, Pankreas-, Blasenkarzinom) ist es auch sehr schwierig, Normalepithel vom gleichen Patienten zu gewinnen. Hier muss man sich mit nicht gepaarten Geweben begnügen bzw. verschiedene Tumorstadien miteinander vergleichen, z.B. Tumorzellen von invasiven mit nichtinvasiven Tumoren oder Primärtumore mit Rezidiven oder Metastasen. Anschließend wird die RNA aus den Tumorzellen bzw. aus den Epithelien in parallelen Ansätzen isoliert, amplifiziert und mit Fluoreszenzfarbstoffen markiert. Dann werden mit Hilfe von proprietären hochdichten Genchips RNA-Profile der Tumorzellen und der korrespondierenden Normalzellen erstellt (Abb. 10). Für jedes Indikationsgebiet werden RNA-Profile von 25-50 Patienten erhoben und über eine bioinformatische Plattform miteinander verglichen. Wir benutzen seit über fünf Jahren die Affymetrix Chip Technologie, da sie erlaubt, eigene hochdichte Genchips zu entwerfen, die sich für die gleichzeitige Analyse von 6.000 bis zu 20.000 Genen sehr gut eignen. Die proprietären metaGen Chips zeichnen sich durch ein hohes Discovery Potential aus, da sie auf die Erfordernisse der Targetidentifizierung der Biotech- und Pharmazeutischen Industrie spezifisch zugeschnitten sind (9). Kandidatengene, die in einer großen Zahl von Tumorpatienten differentiell exprimiert d.h. z.B. hochreguliert sind, werden durch eine Vielzahl von *in-vitro* und *in-vivo* Assays validiert, bevor eine Entscheidung über eine mögliche präklinische Entwicklung einzelner Kandidaten getroffen wird.

## Literatur

- Dunham, I. et al. The DNA sequence of human chromosome 22. *Nature* 1999; 402: 489–495.
- Hattori, M. et al.. The DANN sequence of human chromosome 21. *Nature* 2000; 405: 311–319.
- Lander, E. et al. Initial sequencing and analysis of the human genome. *Nature* 2001; 409: 860–921.
- McPherson, J.D. et al. A physical map of the human genome. *Nature* 2001; 409: 934–941.
- Venter, J.C. et al. The sequence of the human genome. *Science* 2001; 291: 1304–1351.
- Chinwalla A.T. et al. Initial sequencing and comparative analysis of the mouse genome. *Nature* 2002; 420: 520–562.
- Sachidanandam, R. et al. A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. *Nature* 2001; 409: 928–933.

- Futreal, P.A., Kasprzyk, A., Birney, E., Mullikin, J.C., Wooster, R., Stratton, M.R. Cancer and genomics. *Nature* 2001; 409: 850–852.
- Pilarsky, C.P., Schmitt, A.O., Dahl, E., Rosenthal, A. Microarrays – chances and challenges. *Current Opinion in Molecular Therapeutics* 1999; 1: 727-736.

### Tabellen und Abbildungen \*

Sequenzierungs Center	Leiter	Humane Sequenz (Mb)	
		Total	Fertige
1. Whitehead Institute, Cambridge	E. Lander	1196,8	46,5
2. The Sanger Centre, Hixton	J. Sulston	970,7	284,3
3. Washington University, St. Louis	R. Waterston	765,9	175,2
4. DOE JGI, Walnut Creek	T. Hawkins	377,9	78,4
5. Baylor College of Medicine, Houston	R. Gibbs	345,1	53,4
6. RIKEN, Tokyo	Y. Sakaki	203,1	16,9
7. Genoscope, Evry	J. Weissenbach	85,9	48,8
8. Genome Therapeutics, Waltham	D. Smith	71,3	7,0
9. IMB, Jena	A. Rosenthal	49,8	17,7
10. Genomics Institute, Beijing	H. Yang	42,8	6,2
11. Institute of Systems Biology, Seattle	L. Hood	31,2	9,6
12. Genome Technology Centre, Stanford	R. Davies	29,7	3,5
13. Human Genome Centre, Stanford	R. Myers	28,1	9,1
14. University of Washington, Seattle	M. Olson	24,1	14,6
15. Keio University, Tokyo	N. Shimizu	17,3	13,0
16. University of Texas, Dallas	G. Evans	11,6	7,0
17. University of Oklahoma, Oklahoma	B. Roe	10,0	9,1
18. MPI Molekulare Genetik, Berlin	J. Ramser	7,6	2,9
19. GBF, Braunschweig	H. Blöcker	4,6	2,3
20. Cold Spring Harbor Laboratory	R. McCombie	4,3	2,1
Other		59,5	35,9
<b>Total</b>		<b>4338,2</b>	<b>842,0</b>

Tab. 1:  
Humane DNA Sequenz in der HTGS-Abteilung von Genbank im October 2000

---

\* Der Beitrag von André Rosenthal enthält im Original mehrere farbige Abbildungen, deren Erkennbarkeit bei der Graustufen-Wiedergabe beeinträchtigt ist. Interessenten finden die farbigen Abbildungen auf unserer Homepage unter <http://www.leibniz-sozietät.de/sb/58.htm> (Die Redaktion)

<b>Tumor</b>	<b>Neue Fälle in 2001</b>	<b>5-Jahre Überleben (%)</b>
Prostata	198.000	93
Brust (F)	194.000	85
Lunge	170.000	14
Colorectal	135.000	62
Lymphom	64.000	51
Blase	54.000	82
Melanom	51.000	88
Uterus	51.000	84
Leukämie	32.000	43
Niere	31.000	61
Mundraum	30.000	53
Pankreas	29.000	4
Eierstock	23.000	50
<b>Alle Organe</b>	<b>1,270.000</b>	<b>60</b>

*\*Cancer Facts & Figures, the American Cancer Society 2001*

Tab. 2:

*Häufigkeit und 5-Jahresüberlebensraten für die wichtigsten Tumoren in den USA\**

<b>Produkt</b>	<b>Firma</b>	<b>Typ (Target)</b>	<b>Indikation</b>	<b>Jahr der Zulassung</b>	<b>Umsatz/a in Mill \$</b>
Campath	ILEX/ Millenium	humanisiert (anti-CD52)	B-Zell CLL	2001	–
Herceptin	Genentech	humanisiert (anti-ErbB2/Her2)	metastas. Brustkrebs	1998	276
Mylotarg	Celltech/ Wyeth	humanisiert (anti-CD33)	AML	2000	
Rituxan	Genentech	chimär (anti-CD20)	NHL	1997	444
Zevalin	IDEC	Maus (anti-CD20)	B-Zell NHL	2002	–

Tab. 3:

*Zugelassene humane monoklonale therapeutische Antikörper in der Onkologie*

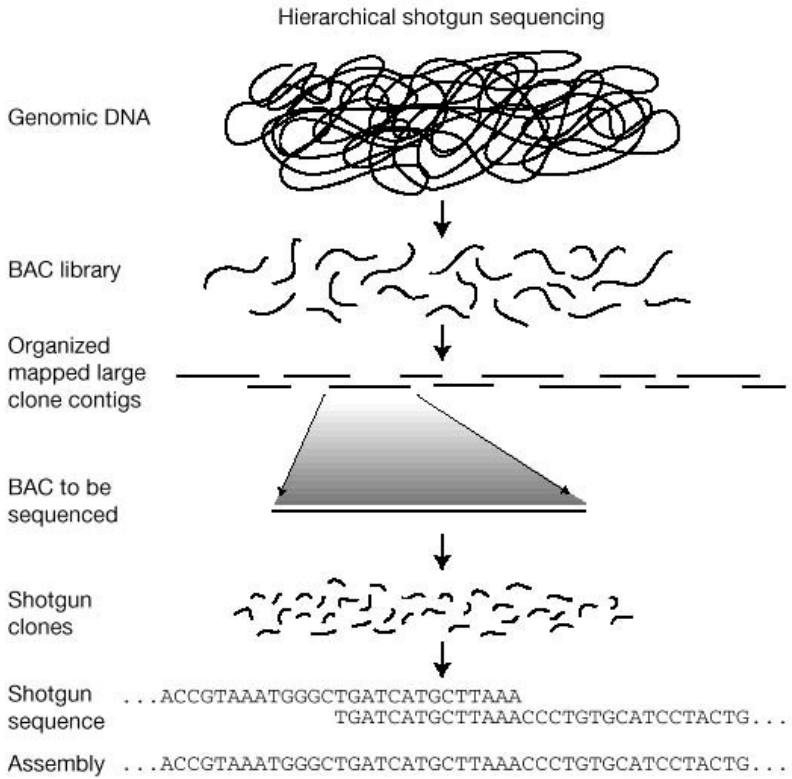


Abb. 1:  
Sequenzierungsstrategie des HGP-Consortiums

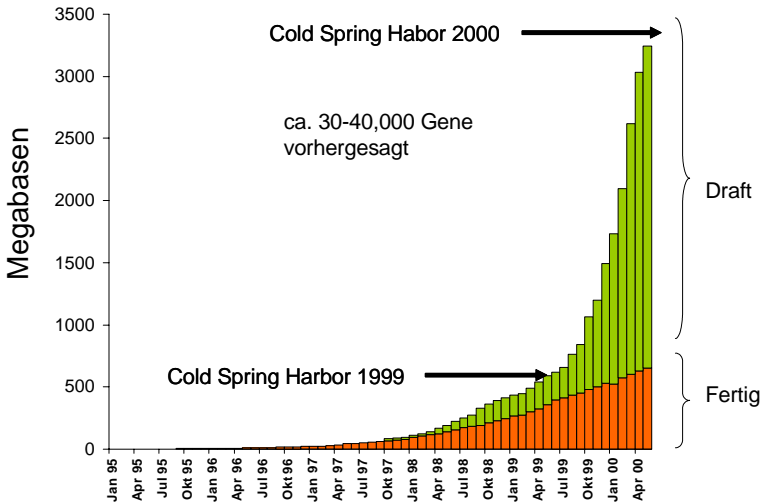


Abb. 2:

Zuwachs der Sequenzen zwischen 1995-2000

- Mehrere tausend Medikamente sind im Markt zugelassen
- Sie basieren aber nur auf ca. 500 bekannten Targets
- Medikamente können sein:
  - kleine Moleküle - Pharma*
  - humane rekombinante Proteine - Biotech's*
  - Monoclonale Antikörper - Biotech's*
  - antisense DANN/ siRNAs - Biotech's*
- Das Genom enthält mehrere tausend neue Targets
- Die Herausforderung besteht, diese Targets zu identifizieren und zu validieren
- Neue Medikamente aus dem Humangenom erreichen Markt *erst ab 2010-15*
- wesentlich erhöhte F&E Kosten müssen aufgebracht werden, um diese Möglichkeiten zu nutzen

Abb. 3:

Humangenom: Hoffnung auf neue Medikamente

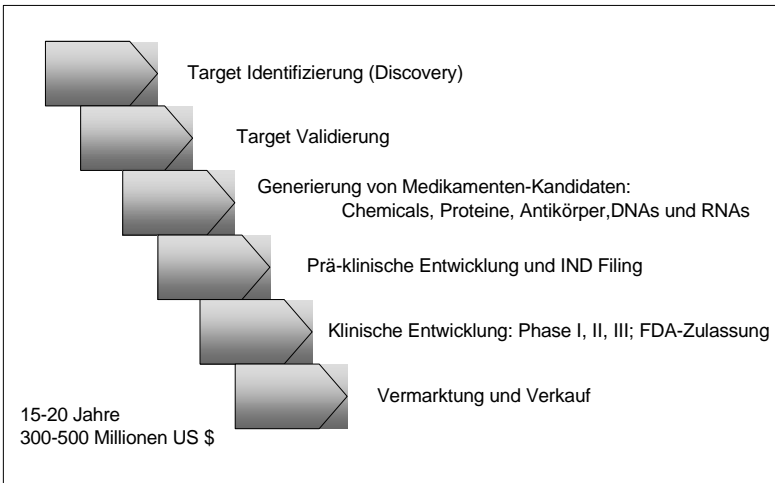


Abb. 4:  
Wertschöpfungskette bei der Medikamenten-Entwicklung

Kategorie	Beispiele
Antimetaboliten	Methotrexat, 5-FU
Monoklonale Antikörper	Herceptin, Zevalin
Mitose Inhibitoren	Taxol
Steroidhormone	Tamoxifen, Flutamid
Alkylierende / vernetzende Medikamente	Endoxan, Cisplatin, Cyclophosphamid
Medikamente, die die Signal-Transduktion modulieren	Gleevec, Tarceva, Iressa
Angiogenese Inhibitoren	Avastin
Telomerase Inhibitoren	BIBR1532
Antitumor Antibiotika	Etoposid, Doxorubicin

Abb. 5:  
Gegenwärtig verwendete Chemotherapien bei Krebs

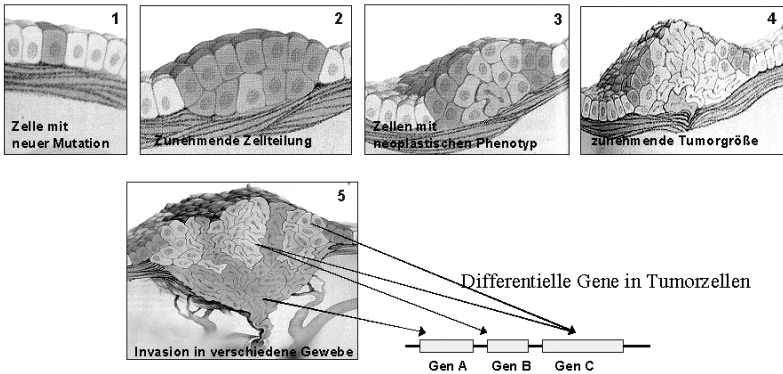


Abb. 6:  
Wie ein Tumor entsteht

Neue Medikamente auf Basis kleiner Moleküle gegen Targets der EGF/EGFR Superfamilie			
Medikament	Firma	klinische Phase	Target
OSI-774 (Tarceva) (HER1)	OSI, Roche, Genentech	III	EGFR
ZD-1839 (Iressa) (HER1)	AstraZeneca	III	EGFR
EKB-569 (HER1)	Wyeth-Ayerst	II	EGFR
PKI-166	Norvartis	I	HER1, HER2
CI-1033	Pfizer	I	HER1-4
GW-2016	GlaxoSmithKline	I	Her1, Her2

Abb. 7:  
Rationales Drug Design bei Krebs



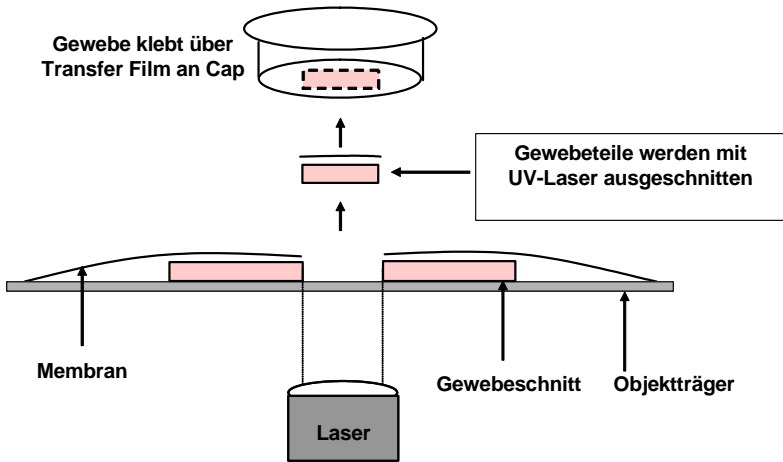


Abb. 8:  
Prinzip der UV-Laser-unterstützten Mikrodisektion

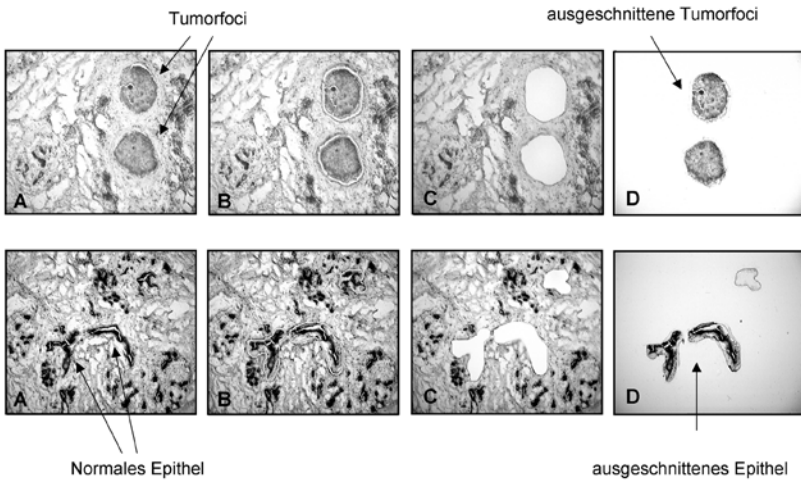


Abb. 9:  
Laser-Mikrodisektion von Brust-Tumoren

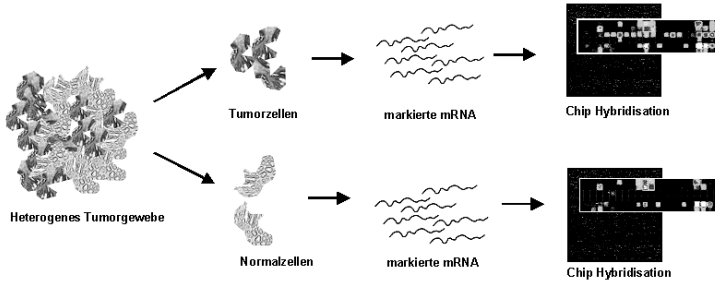
**Detektion von differentiell exprimierten Genen mit Affymetrix Chiptechnologie**

Abb. 10:  
Bestimmung von Tumorprofilen mit Hilfe von Genchips