

Single Nucleotide Polymorphismen (SNPs) – kleine genetische Unterschiede mit großer Wirkung

Dorothee Foernzler

F. Hoffmann-La Roche. Ltd.,
Roche Genetics, Basel

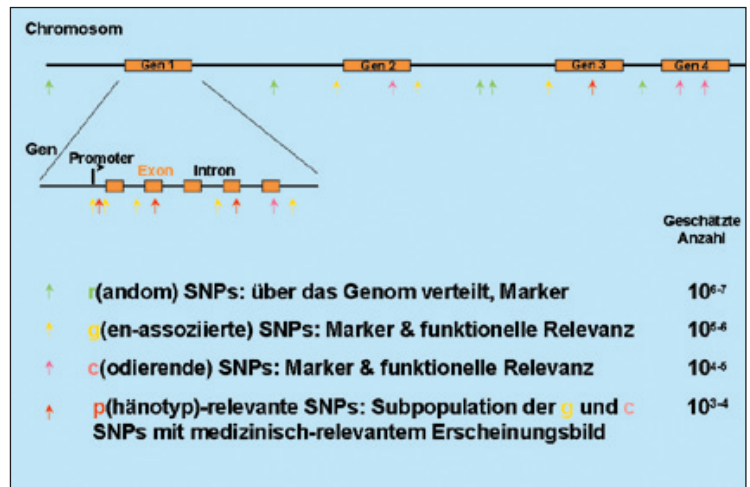


Abb 1 SNP-Klassifizierung aufgrund ihrer Lokalisation im Genom

Zusammenfassung

Kleine genetische Unterschiede in Form von Einzelnukleotidaustauschen im menschlichen Genom, genannt SNPs (Single nucleotide polymorphism), sind zu wichtigen Werkzeugen bei der Erforschung molekularer Mechanismen unterschiedlichster biologischer Funktionen geworden. In der Pharmakogenetik werden SNPs als eine Ursache unterschiedlicher Wirksamkeit von Medikamenten oder für das Auftreten von Nebenwirkungen identifiziert. Bei der Aufklärung von Mechanismen, die zu komplexen Krankheiten führen, sollen SNPs als molekulare Marker helfen, beteiligte Gene zu finden.

Schlüsselwörter
SNP, Pharmakogenetik, Komplexe Krankheiten

Summary

Single nucleotide polymorphisms (SNP), are stable inherited base pair exchanges in the DNA. These polymorphisms are gaining importance in a broad range of applications that try to elucidate the molecular mechanisms underlying different biological functions. In pharmacogenetic approaches SNPs are utilized to explain individual differences in drug response or the occurrence of adverse events. SNP are also used as molecular markers helping to identify genes involved in multigenic inherited diseases.

Keywords

SNP, pharmacogenetics, multigenic inherited diseases

Obwohl seit langem bekannt ist, dass genetische Variationen für die Ausbildung der Blutgruppe, der Haar- und Augenfarbe verantwortlich sind, hat man erst in den letzten Jahren realisiert, dass erbliche Varianten an beinahe allen Aspekten biologischer Funktion beteiligt sind. Daraufhin hat ein wahrer Boom eingesetzt, der sich mit der Erforschung der am häufigsten in unseren Genomen vorkommenden Varianten beschäftigt, den SNPs („Single Nucleotide Polymorphism“). Dieses große Interesse an genetischen Polymorphismen veranschaulicht eine Entwicklung, die sich auf so verschiedene Gebiete wie Populationsgenetik, Forensik, Evolutionsforschung, Pflanzen- und Tierzüchtungsforschung, Medikamentenentwicklung, Krebsforschung sowie der Erforschung komplexer, genetischer Krankheiten ausdehnt. Im folgenden soll die Bedeutung von SNPs bei der Erforschung komplexer, genetischer Krankheiten und bei der Entwicklung und Wirksamkeit von Medikamenten (Pharmakogenetik) näher erläutert werden.

Eigenschaften und Applikationen von SNPs

SNPs sind Einzelnukleotid-Austausche in der DNA, die mit einer Häufigkeit von mindestens 1% in einer Population vertreten sind. Die Attraktivität von SNPs als molekulare Marker ist vor allem auf ihre große Zahl im Genom

Anzeige

**Anforderungen für Assoziationsstudien
(modifiziert nach Emahazion et al., 2001)**

- Selektion der Kandidatengene nach möglicher funktioneller Relevanz für Krankheitsbild oder Medikamentenwirkung (z.B. durch Beteiligung des Gens an bestimmtem Stoffwechsel- oder Signalübertragungsweg)
- Präsenz selektierter SNPs in der zu untersuchenden Population mit statistisch signifikanter Frequenz (abhängig von der Populationsgröße)
- Selektion von SNPs aufgrund ihrer Lokalisation in kodierenden oder regulatorischen Genabschnitten
- Homogene Populationen (Patienten und Kontrollen sollten sich nur in zugrundeliegenden genetischen Risikofaktoren unterscheiden)
- Möglichst große, gut charakterisierte Studienpopulationen
- Replikation der Studie in einer unabhängigen Studienpopulation
- Auswahl geeigneter statistischer Analysemethoden

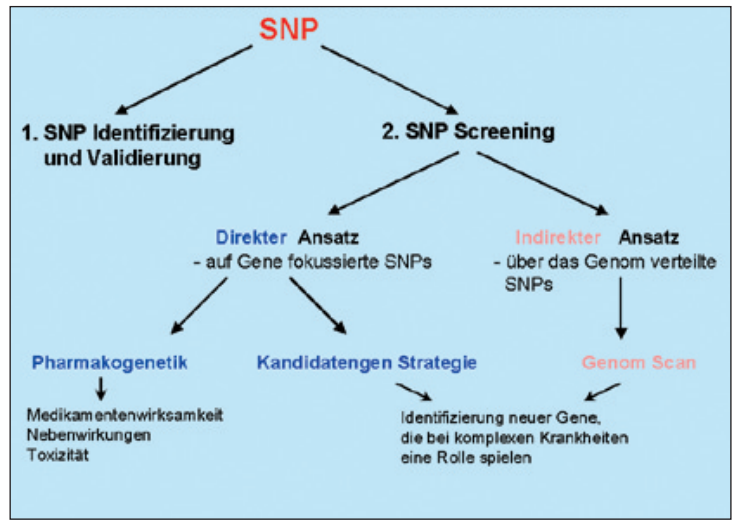


Abb 2 Verschiedene Ansätze von SNP Assoziationsstudien

zurückzuführen. Nach neuesten Schätzungen soll es mehr als 10 Millionen dieser genetischen Varianten geben (Lai, 2001). Darüberhinaus sind SNPs sehr stabil und einfach zu detektieren; nicht zu unterschätzende Eigenschaften, wenn es um Identifizierungs- und Screening-Technologien geht.

SNPs lassen sich aufgrund ihrer Lokalisation im Genom klassifizieren (Abb. 1): Man spricht von sogenannten rSNPs („random SNP“) bei genetischen Polymorphismen, die unabhängig von Genen über das Genom verteilt vorkommen. Diese rSNPs, aus denen sich die SNP Landkarte des menschlichen Genoms hauptsächlich zusammensetzt, werden vorwiegend als molekulare Sonden eingesetzt, mit deren Hilfe man in einem als „Genom Scan“ (siehe unten) bezeichneten „indirekten Ansatz“ den Weg zu Genen bahnen möchte, die bei komplexen, genetischen Krankheiten eine Rolle spielen. (Abb. 2).

Im Gegensatz dazu sind gSNPs („gen-associated SNP“, Abb. 1) mit Genen assoziiert. Unter diesen SNPs genießen vor allem cSNPs („coding SNP“), die in für Aminosäuren kodierenden Genabschnitten lokalisiert sind, und SNPs in regulatorischen Genabschnitten erhöhte Aufmerksamkeit, da sich unter ihnen genetische Polymorphismen mit medizinischer Relevanz befinden (pSNPs, Abb. 1). Zum Beispiel kann ein cSNP einen Aminosäureaustausch im zugehörigen Protein nach sich ziehen. Als Folge davon kann sich die Struktur und/oder die Funktion des Proteins

ändern. SNPs in regulatorischen Genabschnitten können wiederum die Regulation der Genaktivität beeinflussen und damit die Menge des produzierten Proteins verändern. SNPs in Genen können somit – im Gegensatz zu rSNPs – eine „direkte“ funktionelle Relevanz aufweisen und werden in sogenannten „Kandidatengen-Strategien“ (Abb. 2) eingesetzt.

Identifizierung von SNPs im Genom

Die Ermittlung von SNPs erfolgt überwiegend durch Sequenzierung eines bestimmten DNA-Abschnitts in mehreren, nicht verwandten, gesunden Individuen, die verschiedenen ethnischen Gruppen angehören. Daneben werden Polymorphismen in-silico detektiert, d.h durch den Vergleich desselben DNA-Abschnitts aus mehreren Individuen am Computer. Daran sollte sich eine experimentelle Bestätigung des Polymorphismus anschließen (Abb. 3), um zu vermeiden, dass DNA-Sequenzierfehler als genetische Variationen klassifiziert werden. Experimentelle Detektion eines genetischen Polymorphismus erfüllt einen weiteren, für Genotypisierungsstudien wichtigen Zweck: die Validierung des SNPs. Dazu gehören die Bestimmung seiner Häufigkeit, seine Verteilung in verschiedenen Populationen und Angaben darüber, ob eine regulatorische oder kodierende Region eines Gens betroffen ist. Diese Informationen erleichtern die Aufnahme bestimmter SNPs in ein Genotypisierungsprojekt (siehe Textkasten). Es wäre zum Beispiel sinnlos, einen sehr seltenen SNP im Intron eines Gens zu testen, da für

eine solche Variation zu viele Patienten und Kontrollen getestet werden müssten (mehrere 1000 bei einem SNPs von 1-5% Häufigkeit), um zu einer statistisch signifikanten Aussage zu gelangen.

SNP-Landkarte

Mit der Veröffentlichung eines Entwurfs der DNA-Sequenz des menschlichen Genoms ist nun auch erstmals eine SNP-Landkarte mit 1.42 Millionen, im Genom lokalisierten, genetischen Varianten verfügbar (The International SNP Map Working Group, 2001). Fast eine Million dieser Polymorphismen wurden vom sogenannten SNP-Konsortium (TSC) ermittelt, einem einmaligen Zusammenschluss mehrerer Pharma-Firmen und dem Wellcome Trust, das sich die Erstellung dieser genetischen Landkarte zum Ziel gesetzt hatte. Zählt man zusätzliche SNPs aus verschiedenen SNP-Datenbanken und die nicht-redundanten Polymorphismen der kommerziellen SNP-Datenbank von Celera hinzu, sollen bereits über 3 Millionen Einzelnukleotid-Austausche bekannt sein (Lai, 2001).

Diese SNP-Landkarte soll es ermöglichen, Kandidatengene aufzuspüren, die an komplexen, häufigen Krankheiten beteiligt sind. Die meisten häufigen Krankheiten wie zum Beispiel Typ2-Diabetes, Osteoporose oder Alzheimer sind komplexe Krankheiten mit genetischer Komponente. Eines ihrer Charakteristika ist, dass kein einzelnes Gen Risikoträger für die Krankheit ist, wie das für monogene Krankheiten (z.B. Cystischer Fibrose) beschrieben

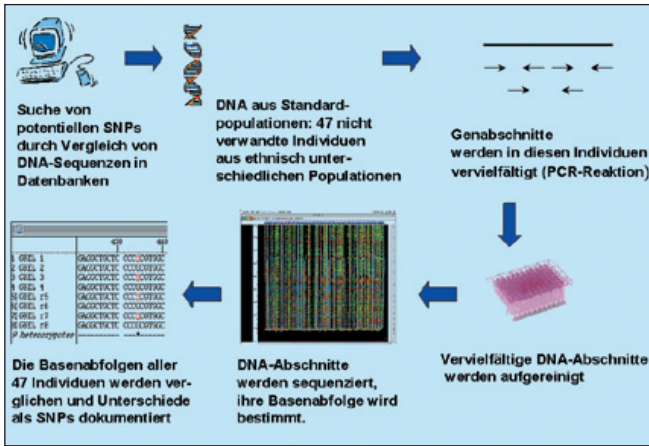


Abb 3 Wie werden SNPs in Kandidatengenen identifiziert?

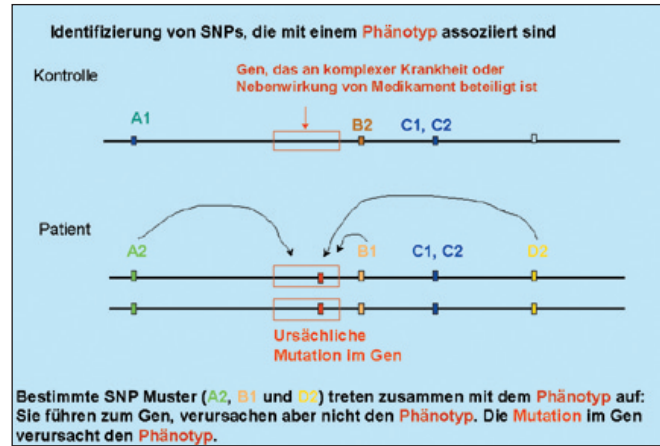


Abb 4 SNPs im „Linkage Disequilibrium“

ist, sondern dass der Krankheitsphänotyp durch die Beteiligung von Defekten in mehreren Genen ausgebildet wird. Da ein einzelnes Gen dabei nur einen moderaten Beitrag leistet, war es bisher sehr schwer, eine Korrelation zwischen diesem Gen und einem Krankheitsphänotyp herzustellen. Dies soll sich nun mit Hilfe von SNPs ändern.

Genom Scan

Die über das Genom verteilten und lokalisierten SNPs der SNP-Landkarte sollen nun in Patienten- und Kontrollpopulationen getestet und miteinander verglichen werden. Es gilt, diejenigen genetischen Polymorphismen herauszufiltern, die den Weg zu in der Nähe liegenden, medizinisch relevanten Genen weisen. Die erfolgreiche Anwendung dieser als „Genom Scan“ bezeichneten Strategie wird zu einem großen Teil vom sogenannten „Linkage-Disequilibrium“ (LD) der SNP Marker abhängen (Abb. 4). LD beruht darauf, dass ein oder mehrere SNPs mit dem Krankheitsphänotyp assoziiert vorkommen. Diese stets miteinander vererbten Genvarianten bilden ein bestimmtes Muster und befinden sich in „Linkage Disequilibrium“. SNPs, die nahe bei einem für die Krankheit relevanten Gen liegen, werden folglich zusammen mit dem Phänotyp vererbt und können aufgrund ihrer räumlichen Nähe auf die genomische Region für ein Kandidatengen hinweisen. Die Anzahl der SNPs für einen „Genom Scan“ hängt dabei vom Ausmaß des LD ab: man vermutet zwischen 3kb (Kruglyak, 1999) und 300kb (Abecasis et al., 2001). Bei einem LD von 3kb

müsste man mindestens 300 000 genetische Varianten testen – und das vermutlich in mehreren 1000 Individuen – eine Herausforderung allein für die Anwendung geeigneter Hochdurchsatz-Genotypisierungs-Technologien und die Finanzierbarkeit solcher Projekte.

Damit werden schon einige Probleme dieses „SNP Genom Scan“ sichtbar: Noch ist unklar, ob die SNP-Dichte der Landkarte ausreicht, um solche Projekte erfolgreich durchzuführen, vor allem unter Berücksichtigung der Tatsache, dass das Ausmaß von LD im Genom je nach Region stark variiert und dass viele genetischen Polymorphismen populationsspezifisch oder zu selten vorkommen. Die Frage wird sein, ob es gelingt, ein für viele Populationen einsetzbares SNP-Set für einen „Genom Scan“ zusammenzustellen.

Kombinierte Strategie

Erfolgreiche SNP-Projekte bei der Identifizierung von Kandidatengenen für komplexe Krankheiten bedienten sich bisher eher einer kombinierten Strategie, bei der mittels klassischer Familienstudien eine genomische Region lokalisiert wurde. In diesen Regionen wurden dann mit Hilfe assoziierter SNPs Kandidatengene identifiziert: Horikawa et al. (2000) detektierten mehrere Polymorphismen im Umfeld des Calpain 10-Gens auf Chromosom 2, die in unterschiedlichen Populationen mit Typ2-Diabetes assoziiert werden konnten. Diese überraschende Entdeckung lässt auf einen neuen, noch unbekanntem Signalweg

zur Entstehung dieser Krankheit schließen. In einer 20 Megabasen umfassenden genomischen Region auf Chromosom 16 haben Hugot et al. (2001) SNP-Muster detektiert, die sie zu NOD2, einem Apoptose-Gen, als Verursacher der Crohn'schen Krankheit geführt haben. Beide Gene hätte man aufgrund ihrer bisher bekannten Funktionen nicht unbedingt in einen „Kandidatengen-Ansatz“ aufgenommen. Der Vorteil eines „Genom Scan“ liegt daher in der Entdeckung von Genen, die bisher nicht in Zusammenhang mit einem bestimmten Krankheitsbild gebracht wurden.

Kandidatengen-Strategie

Die „Kandidatengen-Strategie“ konzentriert sich auf SNPs in selektierten Genen, von denen man eine Beteiligung an einer komplexen Krankheit oder am Auftreten einer Nebenwirkung eines Medikaments bereits vermutet. Der Vorteil dieser Strategie liegt nicht zuletzt in seiner Durchführbarkeit, da in der Regel eine beschränkte Anzahl von SNPs aus Kandidatengenen in charakterisierten Patienten- und Kontrollpopulationen verglichen wird. Die Vorgehensweise ist in Abb. 5 dargestellt. Zur Vorselektion von Genen setzt man oft andere genomische Technologien wie differentielle Genexpression oder Proteomik ein und genotypisiert nur SNPs aus Genen, die entweder unterschiedliche Genexpressions- oder Translationsmuster in Patienten und Kontrollen aufweisen.

Für die Identifizierung von an Osteoporose beteiligten Genen hat man bei-

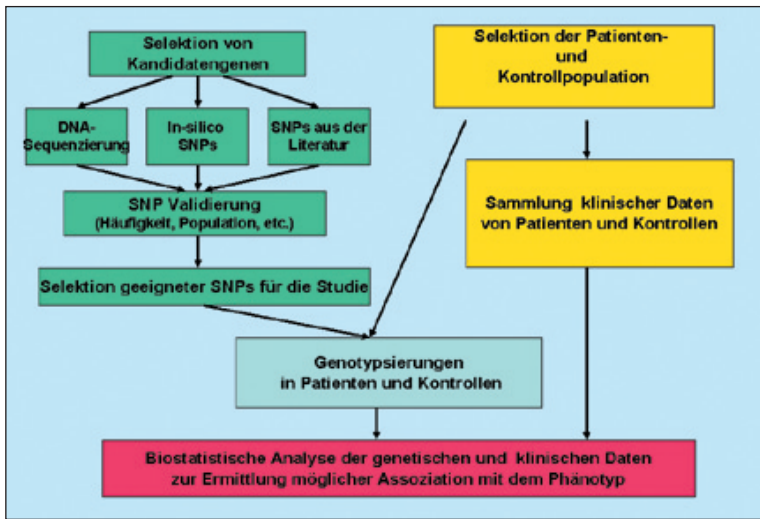


Abb 5 Ablauf einer Assoziationsstudie

spielsweise versucht, SNPs in Genen des Knochen- und Mineralstoffwechsels nachzuweisen, die in einer geeigneten Patienten-/Kontrollpopulation auf Assoziation mit der Knochendichte getestet werden. Für genetische Polymorphismen im Vitamin-D-Rezeptor-Gen wurde eine solche Assoziation nachgewiesen (Alonso et al., 1998).

Neben vielfältiger Anforderungen an Selektion von SNPs und Genen sowie an die zu untersuchenden Populationen (siehe Textkasten) macht auch dieser Ansatz Probleme: Nur wenige der detektierten Assoziationen von SNPs mit einem Krankheitsphänotyp ließen sich mit unabhängigem klinischem Material replizieren. Die Gründe dafür können vielfältig sein: ungenügend charakterisierte oder zu heterogene Testpopulationen, zu kleine Anzahl von Patienten und Kontrollen, populationsspezifische Assoziationen, Wahl der statistischen Analysemethoden, – um nur einige zu nennen (Emaizumi et al., 2001). Man sollte dabei nicht vergessen, dass die Erforschung komplexer Krankheiten mit Hilfe von SNP-Assoziationsstudien noch ein sehr junges Forschungsfeld ist, das Zeit benötigt, um diese Probleme zu lösen. Hoffnungen auf Erfolg machen genetische Varianten, die bereits als Risikofaktoren für komplexe Krankheiten bestätigt worden sind. Das bekannteste Beispiel dürfte ApoE4, eine Variante des Apolipoproteins E-Gens sein, welches für seine Träger mit einem signifikant erhöhten Risiko verbunden ist, an Alzheimer zu erkranken (Roses, 1998).

Einfluss von SNPs auf die Wirkung von Medikamenten

Den bisher größten Einfluss hat die SNP-Kandidatengen-Strategie bisher auf die Pharmakogenetik ausgeübt. Diese Forschungsrichtung beschäftigt sich mit der Tatsache, dass große individuelle Unterschiede sowohl bei der Wirksamkeit eines Medikaments als auch beim Auftreten von Nebenwirkungen bestehen – und das bei fast jedem Medikament (Spear et al., 2001). Beides kann – neben vielfältigen exogenen Ursachen – auf genetische Variationen zurückgeführt werden. Die Kandidatengene für SNP-Studien in der Pharmakogenetik kodieren daher bevorzugt für Rezeptoren, Transporter und andere Medikamenten-Zielproteine sowie für Enzyme, die für die Eliminierung von Medikamenten aus dem Organismus verantwortlich sind. Gerade in letzteren, den Medikamente-metabolisierenden Enzymen, auch DMEs („Drug Metabolizing Enzymes“) genannt, konnten bereits zahlreiche Polymorphismen identifiziert werden, die sich auf die Enzymaktivität und auf eine bestimmte medikamentöse Therapie individuell auswirken. Ein Beispiel ist der Protonenpumpeninhibitor Omeprazol, der zusammen mit dem Antibiotikum Amoxicillin bei der Therapie von bakteriellen Magen- und Darmgeschwüren eingesetzt wird. Omeprazol muss bei Trägern einer bestimmten Genvariante im CYP2C19-Gen für eine erfolgreiche Behandlung höher dosiert werden, da diese Variante Omeprazol zu schnell aus dem Organismus eliminiert (Furuta et al., 2001). Manche SNPs in Zielgenen von Medikamenten lassen be-

reits Aussagen über Nebenwirkungen bestimmter Therapien zu: Bei Asthma-Patienten, die an der Position 17 im beta2-adrenergen Rezeptor ein Arginin anstatt eines Glycins tragen und mit Albuterol behandelt werden, verschlimmern sich die Asthma-Symptome (Israel et al., 2000).

Auch individuelle Reaktionen auf Medikamente können heute schon mit Hilfe von SNPs vorausgesagt werden. Asthma-Patienten mit einer genetischen Variante im Promoter des 5-Lipoxygenase-Gens sprechen nicht auf eine 5-Lipoxygenase-Inhibitor-Therapie an (Drazen et al., 1999). Dies ist ein wichtiger Schritt in die Richtung mehr auf den Patienten zugeschnittener Therapien, in denen ein spezifizierter Einsatz von Medikamenten zum Behandlungserfolg führen und Nebenwirkungen reduziert werden sollen. Darüber hinaus leisten diese SNPs wertvolle Dienste bei der Entwicklung von Medikamenten. Können beispielsweise in einer Phase I- oder Phase II- Studie genetische Varianten identifiziert werden, die mit ungenügender Wirksamkeit assoziiert sind, lassen sich für alle verbleibenden klinischen Studien Probanden selektieren, bei denen das Medikament wirksam ist. Klinische Studien könnten schneller und kostengünstiger durchgeführt werden (Roses, 2000).

SNP-Detektions-Technologien

Die meisten SNP-Applikationen erfordern eine immens große Anzahl von SNP-Genotypisierungstests. Um eine statistisch abgesicherte Assoziation einer Nebenwirkung gegenüber einem

Medikament nachzuweisen, benötigt man mehrere hundert bis mehrere tausend Patienten. Machen mehrere SNPs diesen Phänotyp aus, vervielfacht sich diese Zahl. Dafür benötigt man kostengünstige und automatisierte Testsysteme und entsprechende statistische Analysemethoden, die die Durchführung solcher Mammutprojekte erlauben. Die Biotech-Industrie hat diese Probleme schnell erkannt: Mehr als 40 Firmen wetteifern um die besten Hochdurchsatz-SNP-Screening-Technologien. SNP-Differenzierungs-Methoden beruhen im allgemeinen auf den vier verschiedenen Mechanismen Hybridisierung, Primer Extension, Oligonukleotid-Ligation und strukturspezifischer Spaltung einer Oligonukleotid-Probe. Die Detektion erfolgt in Kombination mit Fluoreszenz-Polarisations- oder Fluoreszenz-Intensitäts-Messungen, Gelelektrophorese oder Massenspektrometrie (Gut, 2001). Die Tendenz führt dabei zur Miniaturisierung des Assay-Formats in den Nanoliterbereich und zur Entwicklung von DNA-Chips, auf denen gleichzeitig mehrere tausend SNPs getestet werden können. Alternativ versucht man, mit Multiplex-Reaktionen oder verschiedenen Pooling-Strategien von DNA-Proben kostengünstiger zu werden. Im Moment scheint aber keine dieser Technologien eine Monopolstellung einnehmen zu können, da die Anwendung einer SNP-Screening-Technologie von der jeweiligen wissenschaftlichen Fragestellung abhängig ist: Ein Projekt erfordert das Testen einer beschränkten Anzahl von SNPs in einer großen Population, ein anderes die Analyse sehr vieler verschiedener SNPs in wenigen Individuen.

Klar ist hingegen, dass der Wert von SNPs darin besteht, sie mit biologischer Funktion und medizinischer Relevanz zu verknüpfen. Halten SNPs was sie versprechen, werden sie in den kommenden Jahren das Verständnis von Krankheitsprozessen und Medikamentenwirksamkeit erheblich erweitern und uns helfen, die medizinische Versorgung zu verbessern.

Literatur

Abecasis GR et al. (2001) Extent and distribution of linkage disequilibrium in three genomic regions. *Am.J.Hum.Genet.* 68:191-197.

Alonso CG et al. (1998). Vitamin D receptor gene (VDR) polymorphisms: effect on bone mass, bone loss and parathyroid hormone regulation. *Nephrol. Dial. Transplant.* 13:73-77.

Drazen JM et al. (1999). Pharmacogenetic association between ALOX5 promoter genotype and the response to anti-asthma treatment. *Nat. Genet.* 22: 168-170.

Emahazion T et al. (2001). SNP association studies in Alzheimer's disease highlight problems for complex disease analysis. *Trends in Genetics* 7: 407-413.

Furuta T et al. (2001) Effect of genotypic differences in CYP2C19 on cure rates for *Helicobacter pylori* infection by triple therapy with a proton pump inhibitor, amoxicillin and clarithromycin. *Clin.Pharmacol.Ther.* 69:158-168.

Gut IG (2001). Automation in genotyping of single nucleotide polymorphisms. *Human Mutation* 17:475-492.

Horikawa Y et al. (2000) Genetic variation in the gene encoding calpain-10 is associated with type 2 diabetes mellitus. *Nat. Genet.* 26:163-175.

Hugot J-P et al. (2001) Association of NOD2 leucine-rich repeat variants with susceptibility to Crohn's disease. *Nature* 411:599-603.

Israel E et al. (2000) The effect of polymorphisms of the beta2-adrenergic receptor on the response to regular use of albuterol in asthma. *Am.J.Respir.Crit.Care Med.* 162:75-80.

Kruglyak L (1999). Prospects for whole-genome linkage disequilibrium mapping of common disease genes. *Nat.Genet.* 22:139-144.

Lai E (2001) Application of SNP technologies in medicine: lessons learned and future challenges. *Genome Res.* 11:927-929.

Roses AD (1998). Apolipoprotein E and Alzheimer's disease. The tip of the susceptibility iceberg. *Ann. NY Acad. Sci.* 855:738-743.

Roses AD (2000). Pharmacogenetics and the practice of medicine. *Nature* 405:857-865.

Spear BB, Heath-Chiozzi M, Huff, J (2001) Clinical applications of pharmacogenetics. *Trends Mol. Med.* 7:201-204.

The International SNP Working Group (2001) A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. *Nature* 409:928-933.

Korrespondenzadresse

Dr. Dorothee Foernzler
F. Hoffmann-La Roche. Ltd.
Roche Genetics, PRBG-T
Bau 93/4.26
CH - 4070 Basel
Tel 0041 61 68 81117
dorothee.foernzler@roche.com