

Zusammenfassung

In den vergangenen Jahren wurden eine Reihe verschiedener Methoden zur Genotypisierung entwickelt, die jeweils ganz eigene Vorzüge, jedoch auch Nachteile aufweisen. Im Folgenden soll in der Hauptsache das Pyrosequencing-Verfahren besprochen werden, das auf dem Prinzip der Sequenzierung durch Synthese beruht und sich als vielseitig einsetzbares Hilfsmittel für die Genotypisierung erwiesen hat. Die dem Verfahrensprinzip inhärente Flexibilität der Versuchsanordnung ermöglicht robustes und genaues Erfassen von Substitutionen wie auch Insertionen/Deletionen. Es eignet sich damit für die Haplotypenanalyse von nahe beieinander liegenden multiplen SNPs ebenso wie für die Multiplex-SNP-Analyse. Da es einen hohen Automatisierungsgrad der Probenpräparation und -analyse ermöglicht, eignet sich das Verfahren auch für die Genotypisierung im großen Maßstab. Im vorliegenden Artikel besprechen wir einige Anwendungen für Pyrosequencing und betrachten mögliche Versuchsanordnungen.

Schlüsselwörter

SNP Typisierung, Pyrosequenzierung; PCR;

Summary

During the last years, a number of different genotyping methods have been developed, each with its different merits and drawbacks. Here we will discuss mainly the pyrosequencing technique, which is based on a sequencing-by-synthesis principle and has been shown to be a flexible tool for genotyping. Through the flexibility in assay design inherent to the principle of the method, robust and accurate scoring of substitutions as well as insertion/deletions is accomplished. It can be used for haplotype analysis of multiple SNPs located proximally, and for multiplex SNP analysis. Together with a high degree of automation in sample preparation and analysis, it is suited for large scale genotyping. In this article, we discuss some applications of pyrosequencing together with assay design considerations.

keywords

SNP typing, pyrosequencing; PCR;

Einführung

Die zunehmende Beachtung der Einzel-Nukleotid-Polymorphismen (SNPs) im Zusammenhang mit einer Reihe komplexer Krankheitsbilder (Risch and Merikangas, 1996) hat zu einem Bedarf an zuverlässigen, für Großanalysen geeigneten Methoden der Genotypisierung geführt. Im Unterschied zur DNA-Sequenzierung, bei der die Didesoxy-basierte Sanger-Sequenzierung als Methode der Wahl gilt, hat sich zur SNP-Analyse noch keine Technologie als weithin anerkannter Standard durchgesetzt. Entscheidende Kriterien für die Akzeptanz einer Methode sind zum einen die Messgenauigkeit, die eine äußerst zuverlässige und praktisch 100% genaue Unterscheidung homozygoter und heterozygoter Allele im diploiden humanen Genom ermöglichen muss, und zum anderen die Flexibilität hinsichtlich der verschiedenen Arten von Polymorphismen, die analysierbar sind (einbezüglich Insertionen/Deletionen). Ebenso wichtig sind die Eignung zur Multiplex-Analyse sowie die Anwendbarkeit zur Bestimmung von Phasen/Haplotypen von SNPs, die auf demselben Gen lokalisiert sind. Die Mehrzahl der Methoden zur SNP-Genotypisierung lassen sich in zwei Gruppen unterteilen: Tests auf Hybridisierungsbasis und enzymatische Diskriminierung. Die Hybridisierung mit Allel-spezifischen Sonden ist die am weitesten verbreitete Methode zur Erkennung von SNPs (Hacia et al., 2000). Allerdings kann die intra- bzw. intermolekulare Struktur der Targets speziell bei Insertionen/Deletionen das Hybridisierungsergebnis unvorherseh-

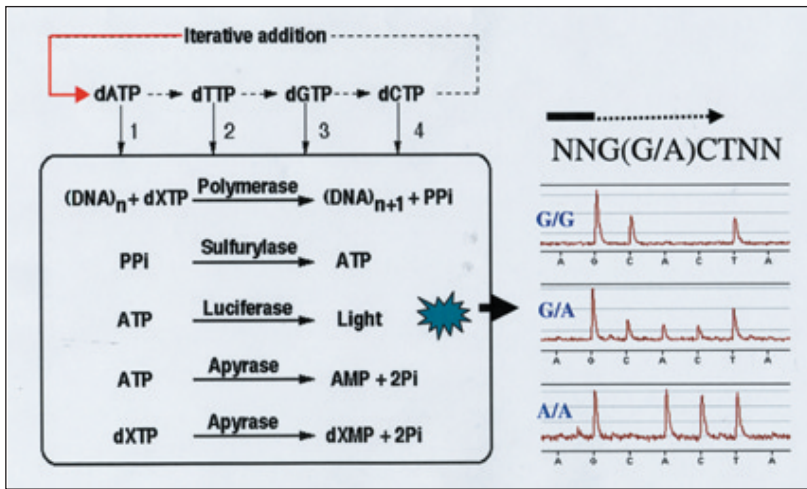


Abb 1 Grundprinzip des Pyrosequencing

Der Einbau eines Nucleotids in eine wachsende DNA-Kette (anfänglich aus einem Primer erweitert) wird durch ein Lichtsignal angezeigt, dessen Intensität proportional zur Anzahl der eingebauten Nucleotide ist. Im Anschluss an die Zugabe eines spezifischen Nucleotids (dXTP) zum Reaktionsgemisch, das neben dem DNA-Templete mit annealem Primer eine Mischung von Enzymen enthält, spalten sich ein Pyrophosphatmolekül (PPI) ab sobald die Erweiterung der Kette durch ein komplementäres dNTP erfolgt. Das PPI wird dann durch die Wirkung der Sulfurylase (im Verbund mit Substrat-AMP) in ATP verwandelt. ATP wiederum ist Substrat für das Licht erzeugende Enzym Luciferase. Gleichzeitig verdaut die Apyrase sowohl nicht eingebaute dXTPs als auch erzeugte ATP, doch hinkt dieser Prozess wegen der geringeren Reaktionskinetik der durch Luciferase vermittelten Lichterzeugung hinterher. Durch iterative Nucleotidzugaben wird die Sequenz der wachsenden DNA-Kette abgeleitet.

bar machen, da solche Targets mit Wildtypsonden unter Umständen bauchig strukturierte Duplexe bilden. Heterozygote Varianten sind im allgemeinen auch schwieriger zu erkennen als homozygote Basenveränderungen, da Hybridisierungssignale von Wildtypallelen das Hybridisierungssignal eines mutierten Allels maskieren und somit das Messergebnis verfälschen könnten. Der Versuch, dies durch die Verwendung von DNA-Chips mit einer Vielzahl redundanter Sonden pro SNP und Referenz-Targetsequenzen zu kompensieren, führte zu sehr komplexen Assays, bei denen es zur Interpretierung der Daten mathematischer Algorithmen bedarf. (Hacia et al., 1996) (Wang et al., 1998). In enzymatischen Extensions-Assays ist die Hitzebeständigkeit des Duplex für die Genotypdiskriminierung nicht entscheidend, da das hybridisierte Target als Template für die weitere Extension durch DNA-Polymerase mit Hilfe markierter Nucleotide verwendet wird. Bei Anwendungen, in denen nur um eine Base verlängert wird, setzt man markierte Didesoxynucleotide ein, die zum Abbruch der Kettenverlängerung führen. Dadurch wird die Bestimmung der Base, die unmittelbar 3' auf den Extension-Primer folgt, möglich. Die Allel-spezifischen Methoden beruhen auf der Extension von Allel-spezifischen Primern, die sich an ihrem 3'-Nucleotid unterscheiden, welches die Allelvarianten der SNPs bestimmt. Ein Problem, das die Entwicklung Allel-basierter Diskriminierungsmethoden behindert hat, ist der Umstand, dass auch mismatched-Primer der Extension unterliegen können. So sind bei

PCR-Reaktionen zum Beispiel GT- bzw. CA-Fehlpaarungen eindeutig nicht von der Extension ausgenommen. (Kwok et al., 1990) (Huang et al., 1992)

Pyrosequencing – Prinzip und Anwendungen

Das Pyrosequencing-Verfahren (Ronahi et al., 1998) ist eine enzymatische „Sequencing-by-Synthesis“-Methode, die sich zur Sequenzierung kurzer DNA-Stränge eignet (bis 50 bp ist eine gute Datenqualität möglich, mündl. Information von Prof. Pål Nyren). Auf Pyrosequencing basierende Versuchsanordnungen zur Genotypisierung lassen sich auf einfache Weise für eine zuverlässige Messgenauigkeit und Typisierung von Substitutionen wie auch von Insertionen/Deletionen (jeweils an späterer Stelle besprochen) entwerfen und sind sowohl für Multiplex-Analyse als auch Haplotypen-Analyse geeignet (Odeberg et al., 2001). Das Grundprinzip ist in Abbildung 1 skizziert.

Das Prinzip der Sequenzierung durch Synthese wird auf die so genannte Out-of-Phase-Extension an verschiedenen polymorphen Allelen angewendet. Durch Änderung der Reihenfolge der Nucleotidaddition erhält man infolge zeitliche Abweichungen in der Allelextension über einige wenige Nucleotidpositionen hinweg, wobei ein Allel zurückbleibt, sehr aussagekräftige Profile für die verschiedenen Genotypen. Dies gewährleistet nicht nur die zuverlässige Interpretation der Rohdaten (präzises „Scoring“) wie in Bild 2 am Beispiel dargestellt, sondern

ermöglicht auch Multiplex-Pyrosequencing, bei dem mehrere SNPs zugleich vervielfacht und analysiert werden. Bei der SNP-Genotypisierung werden Genotypen weniger anhand der relativen Höhe der Peaks an ein oder zwei Positionen unterschieden, sondern vielmehr daran, ob bei einer bestimmten Reihenfolge der Nucleotidaddition (Positionen im Daten-Readout) Peaks auftreten oder nicht. Im Vergleich zur Allel-spezifischen Extension, die ein Ja/Nein-Signal erzeugt, werden hier Muster vom mehreren informativen Peaks generiert, die selbst bei schwachen Signalen verlässliche Messungen zulassen. Somit besteht die Möglichkeit, das Pyrosequencing für die relative Quantifizierung von Virenpopulationen einzusetzen, um das Auftreten sich entziehender Mutanten anhand spezifischer Hot-Spots im Rahmen der medikamentösen Behandlung zu überwachen. Da die Signalstärke zur relativen Target-Menge proportional ist, bedeutet die Einbeziehung mehrerer informativer Peaks in die Analyse einer sich dynamisch verändernden Virenpopulation, dass eine aus der relativen Peak-Höhe an verschiedenen relevanten Positionen abgeleitete Quantifizierung notwendigerweise in einem zuverlässigeren und präziseren Ergebnis mündet (siehe Abbildung 2).

Die Methode eignet sich für die Genotypisierung nicht nur von Substitutionen, sondern auch von Insertionen/Deletionen. Ein entsprechendes Beispiel zeigt Bild 3, wo Out-of-Phase-Extension zu deutlich unterscheidbaren Genotypprofilen führt. Das Verfah-

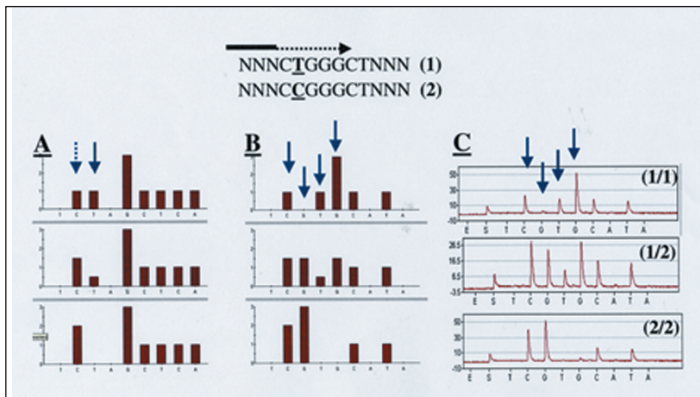


Abb 2 Vergleich von gewöhnlicher bzw. Standardextension und Out-of-Phase-Extension

In A sind die theoretischen/vorhergesagten drei Genotypprofile für die standardmäßige Reihenfolge der Nucleotidaddition dargestellt,

in B die theoretischen/vorhergesagten Genotypprofile, die bei einer optimierten Additionsfolge der Nucleotide, resultierend in der so genannten Out-of-Phase-Extension, entstehen.

C zeigt die tatsächlichen Rohdatenprofile zu letzterem, für die drei verschiedenen Genotypen. Die Extension an den beiden Allelen wird nach der 5. Nucleotidaddition (G) in Phase gebracht. Die Unterscheidung der Genotypen erfolgt durch mehrere informative Peaks, und zwar nicht durch die relative Peak-Höhe an zwei Positionen (wie in A), sondern vielmehr durch das Vorhandensein bzw. die Abwesenheit von Peaks bei bestimmten aufeinanderfolgenden Nucleotidadditionen (Pfeile). Informativen Peaks sind durch Pfeile gekennzeichnet. Wie am deutlichsten in C für den heterozygoten Typ sichtbar, weicht die relative Höhe der beiden ersten C- und G-Peaks von der Vorhersage ab, wobei C ein intensiveres Signal abgibt. Zu einem geringeren Grade ist dies auch beim homozygoten Typ (1/1) erkennbar, wo das erste C im Vergleich zu den darauf folgenden T, C und A größer als vorhergesagt ist. Das Phänomen ist reproduzierbar und sequenzabhängig, und gelegentlich ist es deutlicher ausgeprägt als in diesem Beispiel zu sehen. Die Out-of-Phase-Extension verleiht dem Assay eine ausreichende Robustheit, so dass derartige Abweichungen von den vorhergesagten Profilen die Scoring-Präzision nicht beeinträchtigen.

ren ist indes nicht auf die Genotypisierung kurzer Insertions-/Deletions-Polymorphismen beschränkt. Auch größere bekannte Varianten von Insertionen/Deletionen, die nicht homopolymer sind, lassen sich einfach typisieren. In solchen Fällen bringt man eine anfänglich phasenverschobene Extension nicht wie bei den oben beschriebenen Situationen nach einigen wenigen Nucleotidadditionen in Phase, sondern man bestimmt die Reihenfolge der Nucleotidaddition so, dass **drei sich deutlich voneinander unterscheidende Genotypprofile vorliegen** (siehe Abbildung 3).

Bei enger Nachbarschaft mehrerer SNPs bietet die Methode die Möglichkeit der Phasenbestimmung und Analyse von Haplotypen. Gegenwärtig werden Haplotypen polymorpher Loci, die 30 bp auseinander liegen, bestimmt (Odeberg et al., 2001) und das

Limit liegt nahe der 50 bp-Leseweite, die eine gute Datenqualität bietet.

Das derzeitige Systemdesign ermöglicht die Analyse von 96 Proben in 10 Minuten. Im Verbund mit einer automatischen Probenvorbereitung auf Grundlage der Festphasentechnologie mit Streptavidin-beschichteten Magnet-Beads, wobei Strangseparation und Waschungen unmittelbar in den Pipettenspitzen erfolgen, erreicht die Kapazität eines Instruments 3 000 Genotypanalysen pro Tag. Es steht eine Instrumentenausstattung für parallele Analyse im 384-Well-Format zur Verfügung, und das miniaturisierte Probenformat (10 µl) wird im Verein mit einem Systemdesign, das vom chargenweisen Eintrag von MTP mit PCR-Produkten bis zur Ausgabe der Analysedaten reicht, eine automatisierte, unbeaufsichtigte Analyse mit hohem Durchsatz ermöglichen. Angesichts einer theoretischen Kapazität

von 40 000 Genotypen/Tag wird dieser Engpass bei der Genotypisierung eindeutig auf die vorbereitenden Verfahren verlagert.

Generelle Probleme

Zu den Engpässen der Genotypisierung zählt auch das PCR-Setup, wobei die Automatisierung dieses Schrittes die Entwicklung spezieller, die Kontamination vermeidender Roboter erfordert. Da die Gefahren und Folgen einer Kreuzkontamination zwischen den Kavitäten schwerwiegender sind als die Hochdurchsatzverfahren zur cDNA-Library-Sequenzierung (ESTs), kann man keine Standard-Pipettierroboter einsetzen. Bei nur drei möglichen Genotypen bleibt eine eventuelle Kreuzkontamination zwischen benachbarten Kavitäten bzw. einzelnen DNA-Proben unentdeckt, wenn das gleiche Gen/SNP für viele Individuen auf einer **Mikrotiterplatte** amplifiziert wird. Eine mögliche Strategie, die mit

Anzeige

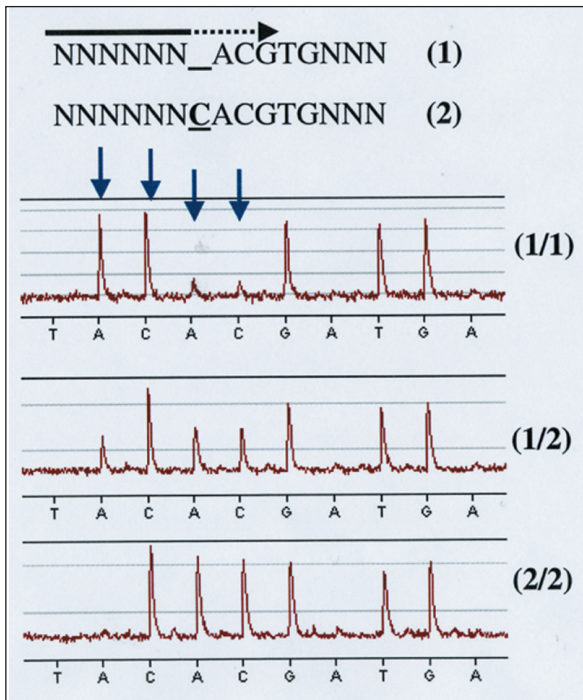


Abb 3 Genotypisierung eines Insertions-/Deletions-Polymorphismus mittels Pyrosequencing bei Anwendung der Out-of-Phase-Extension zur Generierung von drei sich unterscheidenden Genotypprofilen

Pfeile kennzeichnen informative Peaks, die unterschiedliche Genotypprofile anzeigen. Die Extension an Allel (1) eilt bis zur 5. Nucleotid-addition (C) voraus, von welchem Punkt an die Allel-Extension wieder in Phase ist.

üblichen Pipettierrobotern vereinbar ist, besteht darin, verschiedene Genfragmente bzw. SNPs auf einer MTP nur von einem Individuum zu vermehren. In diesem Fall bliebe die Kreuzkontamination zwischen Kavitäten ohne Folgen, so dass nur die Standardmethoden zur Verhinderung der Kontamination von einer MTP zur nächsten erforderlich wären. Dieser Ansatz ist für Genom-weites Scanning geeignet. Ebenfalls von Bedeutung ist die Minimierung der Menge genomischer DNA, die für jede Genotypisierungsoperation bzw. PCR-Reaktion verbraucht wird, da die Biobanksammlungen kostbare und nicht unerschöpfliche Ressourcen darstellen. Eine nested-PCR in Kombination mit Multiplex-PCR-Ansätzen reduziert den Verbrauch an genomischer DNA je Typisierung auf den Sub-Nanogramm-bereich. Nach unserer Erfahrung ist die Einrichtung von Multiplex-Ansätzen, auf denen 10-12 Gene bzw. SNPs gemeinsam vervielfacht werden, recht effizient, sofern man einige Grundregeln des Primer-Designs befolgt: hohe Annealing-Temperaturen (68–72 °C), Primer-Länge unterhalb 25 bp, Länge des Target-Fragments unterhalb 500 bp. Obwohl die theoretische Möglichkeit besteht, für eine sehr große Zahl von Genen Multiplex-Ansätze einzurichten, erscheint der Optimierungsaufwand zur Entwicklung von Panels für die simultane Vervielfachung von 20 bis 25 Genen nicht unbedingt sinnvoll, wenn Gen-spezifische Primer Anwendung finden. Eine Strategie zur Vervielfältigung ganzer Genome wäre hingegen eine attraktive Alternative zur Gen-spezifischen

Multiplex-Vervielfältigung **in der äußeren PCR**, wenn auch die Robustheit und Zuverlässigkeit der Methoden zur Vervielfältigung des ganzen Genoms noch weiterer Entwicklung und Auswertung bedarf (Wells et al., 1999).

Vergleich zu anderen derzeit genutzten Technologien

Die Genotypisierungsverfahren unter Verwendung kommerzieller Systeme auf Grundlage der MALDI-TOF Massenspektrometrie, beispielsweise die Sequenom-Technologie, weisen bis zum eigentlichen Analyseschritt (PCR-Vervielfältigung, Festphasen-Probenvorbereitung von DNA-Einzelsträngen und Primer-Annealing für die anschließende enzymatische Extension) ein recht ähnliches Systemdesign auf und unterscheiden sich im Grunde erst hinsichtlich der Analyse als solcher. Zu den Vorteilen auf MALDI-TOF MS basierender Methoden zählen die geringen Probenmengen, die für den Nachweis erforderlich sind, und ihre Eignung zur Typisierung von Tandem-Wiederholungen und längeren homopolymerischen Insertionen/Deletionen (> 6–10 bp), also Genomvariationen, die sich durch Pyrosequencing schwer analysieren lassen. Zu den Nachteilen gehören die relativ kostenaufwendige Geräteausstattung, die Vielzahl von (wenn auch automatisierten) Analyseschritten sowie bei der Mehrzahl der auf MALDI-TOF beruhenden Methoden die äußerst hohen Anforderungen an die Reinheit des zu analysierenden Produkts. Pyrosequencing und MALDI-TOF MS sind hinsichtlich des hohen Probendurch-

satzes und Automatisierungsgrads durchaus vergleichbar, wobei das Pyrosequencing-Verfahren so flexibel ist, dass es auch auf Genotypisierungsvorhaben in kleinen und mittelgroßen Labors anwendbar ist. Die Vorteile von Pyrosequencing liegen neben den geringeren Gerätekosten in der Fähigkeit, problemlos eine Haplotypanalyse eng benachbarter SNPs (wie bereits oben erwähnt) sowie die Typisierung langer Insertionen/Deletionen, die weder homopolymerisch noch Di-/Tri-Nucleotid-Wiederholungen sind, zu ermöglichen. Der von Gut und Mitarbeitern entwickelte GOOD-Assay umgeht viele Beschränkungen anderer MALDI-TOF-basierter Methoden und ermöglicht die Verwendung von ungereinigtem PCR-Produkt als Proben (Sauer et al, 2000). Wie uns berichtet wurde, ist es zur Bestimmung von Haplotypen in der Lage (Dr. Ivo Gut, mündliche Information) und zählt zu den eher vielversprechenden Genotypisierungsmethoden auf MALDI-TOF-Grundlage, obwohl die Gerätekosten wohl nur Großlabors den Einsatz dieses Verfahrens gestatten. DNA-Chip-Technologien auf der Basis immobilisierter SNP-spezifischer Primer/Sonden lassen sich auf Genom-weite Scans anwenden, bei denen für jedes Individuum eine sehr große Zahl von SNPs zu analysieren ist (zum Beispiel Linkage-Studien), eignet sich jedoch nicht (bzw. ist zu teuer) für typische Assoziationsstudien, in denen nur eine begrenzte Zahl in Frage kommender SNPs bei einer sehr großen Zahl von Individuen untersucht werden (wofür wiederum MALDI-TOF- und Pyrosequencing-Verfahren sehr geeignet

sind). Darüber hinaus bedürfen Assay- und Probenaufbau noch einiger Optimierung; eventuell wird es sich sogar als unmöglich erweisen, Versuchsanordnungen für SNPs an diffizilen Genorten zu entwickeln. Der zuletzt angeführte Nachteil betrifft auch die TaqMan-Assays, doch haben diese Verfahren wiederum den Vorteil, dass Vervielfältigung und Genotypanalyse hier in einem einzigen Schritt erfolgen. Bei bestimmten SNPs wurden im Zusammenhang mit TaqMan Probleme hinsichtlich einer gesicherten Messgenauigkeit berichtet (Täpp et al, 2000), die beim Pyrosequencing – unter anderem aufgrund der hochcharakteristischen Genotypprofile, die sich aus der Anwendung des Out-of-Phase-Prinzips (wie weiter oben besprochen) ergeben – nicht auftreten.

Schlussbemerkungen

In dem Maße, wie die Genotypisierung von SNPs in großen Umfängen auf verschiedenen Gebieten der Medizinforschung an Bedeutung gewinnt, wird die Forderung nach robusten **Scoring-Systemen** die verschiedenen Kandidaten auf die Probe stellen. Das Pyrosequencing zählt hierbei zu den viel versprechenden Methoden, da diesem Verfahren im Hinblick auf mögliche Versuchsansätze und Analysearten eine hohe Flexibilität zu eigen ist und es sich in automatisierte Systeme mit hohem Analysedurchsatz integrieren lässt.

Übersetzung: Klaus Goldammer

Literatur

Hacia, J. G., Brody, L. C., Chee, M. S., Fodor, S. P. und Collins, F. S. (1996). Detection of heterozygous mutations in BRCA1 using high density oligonucleotide arrays and two-colour fluorescence analysis. *Nat Genet* 14, 441-7.

Hacia, J. G., Edgemon, K., Fang, N., Mayer, R. A., Sudano, D., Hunt, N. und Collins, F. S. (2000). Oligonucleotide microarray based detection of repetitive sequence changes. *Hum Mutat* 16, 354-63.

Huang, M. M., Arnheim, N. und Goodman, M. F. (1992). Extension of base mispairs by Taq DNA polymerase: implications for single nucleotide discrimination in PCR. *Nucleic Acids Res* 20, 4567-73.

Kwok, S., Kellogg, D. E., McKinney, N., Spasic, D., Goda, L., Levenson, C. und Sninsky, J. J. (1990). Effects of primer-template mismatches on the polymerase chain reaction: human immunodeficiency virus type 1 model studies. *Nucleic Acids Res* 18, 999-1005.

Odeberg, J., Holmberg, K., Eriksson, P. und Uhlén, M. (2001). Haplotype analysis by Pyrosequencing. *Eingereicht*.

Odeberg, J., Holmberg, K., and Uhlén, M. (2001). Multiplex analysis by pyrosequencing. *Eingereicht*.

Risch, N. und Merikangas, K. (1996). The future of genetic studies of complex human diseases. *Science* 273, 1516-7.

Ronaghi, M., Uhlen, M. und Nyren, P. (1998). A sequencing method based on real-time pyrophosphate. *Science* 281, 363.

Sauer, S.; Lechner, D.; Berlin, K.; Plancon, C.; Heuermann, A.; Lehrach, H.; und Gut, IG. (2000). Full flexibility genotyping of single nucleotide polymorphisms by the GOOD assay. *Nucleic Acids Res* 28 (23) e100.

Täpp, I., Malmberg, L., Rennel, E., Wik, M. und Syvänen, A-C. (2000). Homogenous scoring of single nucleotide polymorphisms: Comparison of the 5' nuclease TaqMan assay and molecular beacon probes. *Biotechniques* 28, 732-738.

Wang, D. G., Fan, J. B., Siao, C. J., Berno, A., Young, P., Sapolsky, R., Ghandour, G., Perkins, N., Winchester, E., Spencer, J., Kruglyak, L., Stein, L., Hsie, L., Topaloglou, T., Hubbell, E., Robinson, E., Mittmann, M., Morris, M. S., Shen, N., Kilburn, D., Rioux, J., Nusbaum, C., Rozen, S., Hudson, T. J., Lander, E. S. et al. (1998). Large-scale identification, mapping, and genotyping of single-nucleotide polymorphisms in the human genome. *Science* 280, 1077-82.

Wells, D., Sherlock, JK., Handyside, AH. und Delhany, JDA. (1999). Detailed chromosomal and molecular genetic analysis of single cells by whole genome amplification and comparative genomic hybridisation. *Nucleic. Acids Res.* 27, 1214-1217.

Korrespondenzadresse

Dr. Jacob Odeberg
Royal Institute of Technology
Dept. of Biotechnology
100 44 Stockholm, Schweden
Tel. 0046 8 790 8287
Fax 0046 8 24 54 52
jacob@biochem.kth.se