

Genome-wide screening for genetic variations associated with aggressive prostate cancer risk

Identifizierung genetischer Marker für eine aggressive Form des Prostatakarzinoms

*Ao. Univ. Prof.ⁱⁿ Mag.^a Dr.ⁱⁿ Andrea Gsur, Ao.Univ. Prof. Dr. Michael Krainer,
Prof. Thomas Meitinger*

Ausgangssituation

Eines der Hauptprobleme im Management des Prostatakarzinoms ist das Fehlen verlässlicher genetischer Marker, die den weiteren klinischen Verlauf zum Zeitpunkt der Diagnose voraussagen. Obwohl das Prostatakarzinom meist ein relativ langsam wachsender Tumor ist, existieren auch sehr aggressive Formen welche rasch zur Progression und letztendlich zum Tod führen können. Es ist weitgehend ungeklärt welche molekulare Mechanismen für die Progression des Prostatakarzinoms bis hin zur letalen Form verantwortlich sind. Durch die Identifizierung genetischer Risikomarker, die die Differenzierung in eine latente oder aggressivere Form der Erkrankung zum Zeitpunkt der Diagnose ermöglichen könnten Behandlungen (Operation), die zu Inkontinenz und Impotenz führen können und bei einer latenten Form des Prostatakarzinoms nicht notwendig wären, vermieden werden.

Ziele / Methoden

Die Ziele dieses Forschungsprojektes sind einerseits die Identifizierung genetischer Marker, die mit einem höheren Prostatakarzinomrisiko assoziiert sind und andererseits Marker, die zum Zeitpunkt der Diagnose die Differenzierung erlauben, welche Männer eine aggressive Form des Prostatakarzinoms entwickeln werden.

Probenselektion und Qualitätskontrolle für SNP Array 6.0

Insgesamt wurden 250 DNA-Proben aus unserer DNA-Bank ausgewählt. Davon sind 84 Prostatakarzinome mit einem Gleason Score größer 7, 84 Prostatakarzinome mit Gleason Score kleiner 7 sowie 82 benigne Prostatahyperplasien (BPH) als Kontrollgruppe. Bei der Auswahl wurde darauf Wert gelegt, dass die Durchschnittsalter in den drei Gruppen sehr ähnlich sind, sowie wichtige Confounding-Daten wie Angaben zum Rauchverhalten oder der PSA-Wert (Prostata-spezifisches Antigen) von den ausgewählten Datensätzen vollständig sind. Da für die Prozessierung der DNAs vor der Arrayhybridisierung eine DNA-Konzentration von 40–90 ng/µL erforderlich war, wurde die Probenauswahl auf jene DNAs mit Konzentrationen größer 50 ng/µL eingeschränkt. Die DNA-Konzentrationen aller ausgewählten Proben wurden mittels NanoDrop 1000 Spectrophotometer bestimmt und 50 ng/µL-Verdünnungen in Aqua bidest. hergestellt. Anschließend wurden die Konzentrationen der Verdünnungen nochmals überprüft

sowie die Proben verblindet. Da für die DNA-Chips nur solche DNAs eingesetzt werden dürfen, die eine hohe Qualität besitzen und nicht degradiert sind, wurde die DNA-Qualität aller Proben in 0,8%igen Ethidiumbromid-gefärbten Agarosegelen überprüft. Es wurde jeweils 1 µL der 50 ng/µL-Verdünnung eingesetzt und die Dauer der elektrophoretischen Auftrennung betrug 65 min bei 120 Volt. Bei qualitativ hochwertiger (nicht degradierter) DNA ist auf dem Agarosegel eine Einzelbande zwischen 10 und 20 kbp (Kilobasenpaare) zu erwarten während Banden unterhalb dieses Bereichs oder das Auftreten eines "Schmiers" auf beschädigte, abgebaute DNA schließen lassen. Alle Proben zeigten eine hohe Qualität (starke Einzelbande bei 20 kbp).

Genome-Wide Human SNP Array 6.0

Hierbei handelt es sich um einen sogenannten „DNA-Chip“, mit welchem über 1,8 Millionen genetische Varianten einer Person bestimmt werden können. 906.600 dieser genetischen Varianten sind sogenannte „Single Nucleotide Polymorphismen“ (SNPs), welche die häufigste Sequenzvariation im menschlichen Genom bilden und sich wegen ihrer Häufigkeit und Stabilität besonders als Marker für genomweite Assoziationsstudien eignen. Zudem trägt dieser Array 946.000 Sonden zur Bestimmung von „Copy Number“-Variationen (CNVs) wodurch genetische Veränderungen wie Duplikationen, Deletionen oder Inversionen nachgewiesen werden können.

Genomische DNA musste dazu zunächst aufwendig prozessiert werden bis die Proben schließlich auf die Arrays geladen werden konnten und die generierten Genotypisierungsdaten mit einer bestimmten Software ausgewertet werden können. Die DNA wurde eingangs mit den Restriktionsenzymen Sty und Nsp verdaut um in einem nächsten Schritt mittels T4 DNA Ligase an einen Adapter ligiert zu werden. Darauf folgte unter der Verwendung von TITANIUM™ Tag DNA Polymerase eine Vermehrung mittels Polymerase-Kettenreaktion (PCR), deren Produkte anschließend über Filter aufgereinigt wurden. Schließlich erfolgte noch eine Fragmentierung der PCR-Produkte und ihre Markierung mithilfe des Enzyms terminaler Desoxynucleotidyltransferase. Danach wurden die Proben auf die Arrays geladen. Es folgte eine 18stündige Hybridisierung sowie ein Wasch- und Färbeschritt. Danach erfolgte die Aufzeichnung der Rohdaten durch den GeneChip Scanner 3000 7G (Scanning). Als Ergebnis liegen die Daten in *.CEL-Files vor, welche im Anschluss mit der Affymetrix Genotyping Console Software analysiert werden müssen.

Ergebnisse

Im Zuge der Array-Prozessierung mussten 18 der 250 genomischen DNA-Proben wiederholt werden, da es zu einem Problem mit einer PCR-Aufreinigungsplatte kam. Nach erfolgreicher Wiederholung der Platte war die Prozessierung aller 250 Proben erfolgreich. Alle Array-Rohdaten liegen nunmehr in Form von Intensitätsfiles (CEL-Files) vor. Abbildung 1 bietet einen Eindruck eines solchen gescannten „DNA-Chips“. Dabei befinden sich die Sonden zur Bestimmung der SNP-Genotypen in den vier Qua-

dranten, während die Sonden zur Detektion der Copy Number Variations (CNV) im Kreuz zwischen den Quadranten angeordnet sind.

In einem ersten Schritt wurden für die 250 Intensitätsfiles Command Console Sample-Files (ARR) mit allen relevanten klinischen Daten und Sample-Attributen erstellt. Diese ARR-Files wurden über das AGCC-Portal von Affymetrix mittels „Batch Registration“ generiert und gemeinsam mit den Intensitätsdaten in die Genotyping Console 4.0 eingespielt. Zu Beginn der Primäranalyse wurde zur Überprüfung der Datenqualität der CEL-Files eine sogenannte „Contrast QC“ für jede Probe berechnet, welche eine spezielle Testgröße in Genotypisierungsstudien darstellt. Diese Messgröße beschreibt die Tauglichkeit eines Experiments, die aufgelösten SNP-Signale in (die bei SNPs möglichen) drei Genotypcluster zu separieren. Genome-wide SNP_6 verwendet dazu ein statisches Set aus 10.000 randomisiert ausgewählten SNPs und berechnet die Unterschiede in den „Kontrastverteilungen“ zwischen den Signalpeaks der homozygoten und heterozygoten Genotypen (Abbildung 2). Für die folgende Genotypenberechnung mittels „Birdseed“-Algorithmus ist ein Contrast QC-Schwellenwert von 0,4 erforderlich. Alle Proben lagen über diesem Schwellenwert und der Durchschnitt lag bei 2,3, was für eine hohe Güte der Intensitätsdaten spricht (Abbildung 3). Weiters wurde für jede Probe unter Verwendung der Kontrastdaten von allen SNPs des X-Chromosoms eine „Gender Estimation“ durchgeführt. Dabei konnten mit den Default-Einstellungen 248 der 250 Proben eindeutig einem Geschlecht zugeordnet werden.

Die Qualitätsanforderungen der Genotypisierungsberechnung (90% der Proben Contrast QC > 0,4 und durchschnittliche Contrast QC > 1,7) wurden erfüllt (siehe oben sowie Abbildung 4). Dies beschreibt ebenfalls die hohe Güte der Genotypisierungsdaten

Ausblick

Bei der Erstellung von SNP-Listen und SNP Cluster Graphs zur visuellen Kontrolle von ausgewählten SNPs ist derzeit die Rechenleistung unseres Rechners zu gering, weshalb die weitere Auswertung erst nach Eintreffen des neuen Rechners fortgesetzt werden kann. Der nächste Schritt ist das SNP-Filtering, um eine möglichst kondensierte SNP-Liste für die sekundäre Datenanalyse und die Re-genotypisierungsphase zu erhalten. Diese SNPs werden mittels Open Array-Technologie in der gesamten Studienpopulation (ca. 1500 Proben) genotypisiert werden.

Der große Umfang an generierten Genotypisierungsdaten erfordert eine tiefgehende Analyse der signifikantesten SNP-Loci und CNVs. Danach können Entscheidungen getroffen werden, welche Gene, Loci oder Signaltransduktionswege, die das Prostatakarzinomrisiko beeinflussen könnten, noch im Detail untersucht werden sollen um schlussendlich ein Risikoprofil aus einer Kombination vieler genetischer Biomarker für das Prostatakarzinom zu entwickeln.

Diese Forschungsprojekt wird einerseits mit Mittel der „Initiative Krebsforschung“ und größtenteils mit Mitteln des „Bundesministeriums für Gesundheit“ („Identifizierung genetischer Risikofaktoren zur Früherkennung und Prognose des Prostatakarzinoms mittels SNP microarrays“) finanziert.

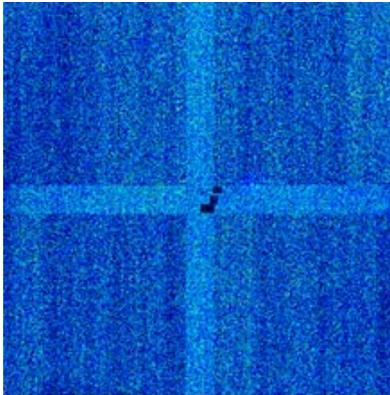


Abb. 1: Genome-wide Human SNP 6.0 Array mit über 1,8 Mio genetischen Markern (906.600 SNPs und 946.000 CNV-Sonden).

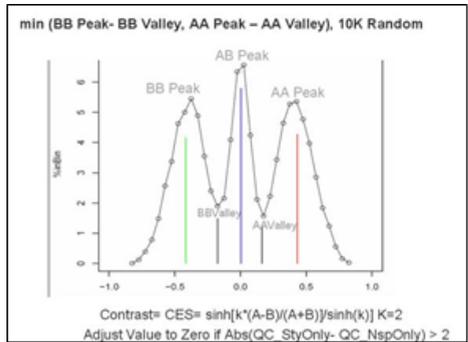


Abb. 2: Berechnung der Contrast QC zur Bewertung der Datenqualität.

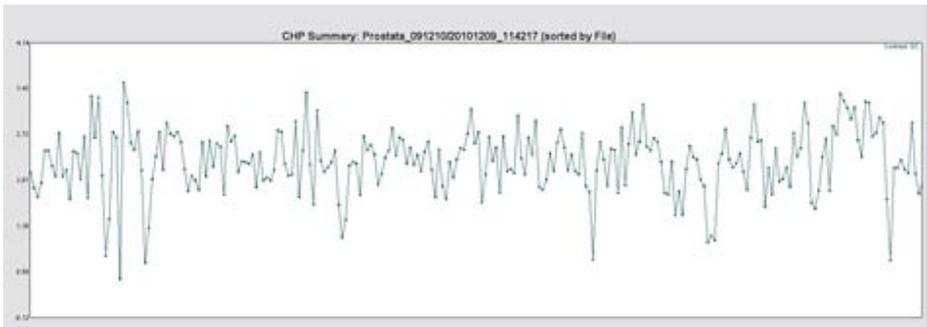


Abb. 3: Contrast QC-Werte aller 250 Arrays (Cut-off 0,4)

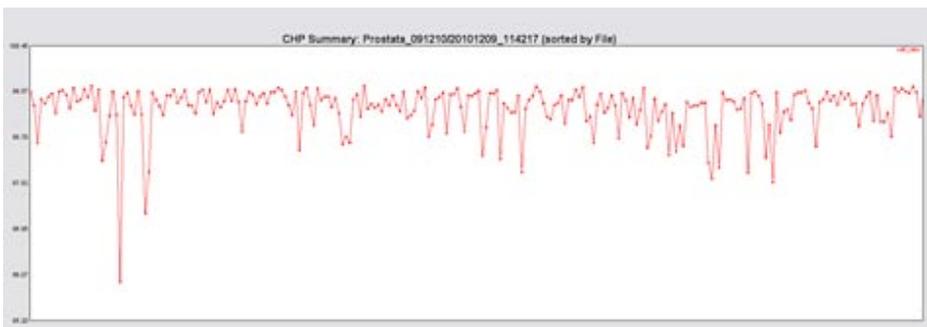


Abb. 4: SNP Call Rates