

Jia-Lu Qiao

Dr. med.

Copy Number Variation in patients with ischemic stroke

Fach/ Einrichtung: Chirurgie

Doktorvater: Priv.-Doz. Dr. med. Philipp Erhart

Die meisten derzeitigen Algorithmen zum Nachweis von Copy Number Variations (CNVs) sind arbeitsintensiv, zeitaufwendig und weisen eine hohe Anzahl von falsch-positiven Ergebnissen auf. Um die Genauigkeit und Geschwindigkeit der Genotypisierung von CNVs mittels Micorarray-Technologie zu verbessern, präsentiert diese Arbeit eine alternative Strategie, bei der die Scatterplot-Methode angewendet wurde. Diese ermöglicht eine Hochdurchsatz-Genotypisierung von CNVs in vordefinierten Regionen großer Kohortenstudien.

Um die Scatterplots zu generieren, musste zunächst eine Liste bekannter CNV-Regionen für die Analyse erstellt werden. Anschließend wurden folgende Schritte sequenziell für alle Proben angewendet: 1) Berechnung der mittleren oder durchschnittlichen Signalintensitäten für alle Einzel-Nukleotid-Polymorphismen (SNPs) innerhalb jeder CNV-Region und ihrer flankierenden Regionen; 2) Normalisierung durch Subtraktion der flankierenden Signale der CNV-Signale und; 3) Kategorisierung aller SNPs in der CNV-Region als homozygot, heterozygot di-allelisch oder heterozygot tri-allelisch mit anschließender Auszählung. Schließlich wurden alle vom Scatterplot-Methode identifizierten CNV-Ergebnisse visuell überprüft.

Es wurden die Python-Codes zur Metrikberechnung und Generierung von Diagrammen bereitgestellt.

Anschließend wurde die dreistufige Vorgehensweise zum Auslesen eines typischen Diagramms demonstriert. Die Verwendung eines Satzes von acht Scatterplots, die CNVs aus verschiedenen Aspekten analysieren, bietet den Vorteil, nicht nur die Genotypisierungsergebnisse effizient darzustellen, sondern auch eine schnelle Bewertung der Datenqualität der Region. Gleichzeitig wurden drei vorteilhafte Eigenschaften des Scatterplots vorgestellt: 1) Die Signal-Korrelation eines CNV und seiner flankierenden Regionen zur Unterscheidung von falsch Positiven, 2) die Normalisierung zur Reduzierung von Störartefakten und 3) die Metrik der B-Allel-Frequenz zur Verbesserung der Genotypisierungsgenauigkeit.

Die Scatterplot-Methodik ist nützlich, um vordefinierte CNVs zu analysieren. Die syndromischen CNVs aus der Studie von Crawford wurden auf die Stichproben der SLESS-Studie angewendet (eine der Kohorten von CaNVAS). Insgesamt trugen 41 Proben 10 verschiedene CNVs. Alle Ergebnisse wurden einer visuellen Validierung unterzogen, wobei die "Noise-free-CNV" Software zum Einsatz kam. Alle Ergebnisse waren korrekt und es gab keine fehlenden Ergebnisse.

Die Robustheit der Scatterplot-Methode für verschiedene Mikroarray-Plattformen wurde bestätigt. Wir führten CNV-Genotypisierungen auf Chromosom 6:31.36-31.45 Megabase und Chromosom 12:7.87-8.02 Megabase in neun Datensätzen durch, die von verschiedenen Mikroarray-Chips generiert wurden. Die durchschnittliche SNP-Dichte zwischen den beiden Regionen und sogar die SNP-Dichten zwischen verschiedenen Chips innerhalb einer Region waren signifikant unterschiedlich. Durch die Scatterplot Methode wurden die Proben deutlich nach ihrem CNV-Status gruppiert, und die Grenzen zwischen den Clustern waren deutlich erkennbar. Die Genauigkeit der CNV-Identifizierung war robust und kongruent.

Das Scatterplot-Tool bietet einen potentiellen Vorteil, strenge QC-Filterungen zu umgehen. In der SLESS Genotypisierung beispielsweise, mussten bei Verwendung der PennCNV-Methode 163 minderwertige Proben (10,27%) ausgeschlossen werden. Dieser Schritt war jedoch bei Verwendung der Scatterplot-Methode nicht erforderlich.

Die Scatterplot-Methode und die traditionelle Methode wurden verwendet, um 692 autosomale CNV-Regionen zu analysieren und deren Effizienz zu vergleichen. Bei der Analyse von Proben des SLESS-Datensatzes (n=1,587) benötigte die traditionelle Methode 2,5-mal mehr Zeit als die Scatterplot Methode. Bei der Analyse größerer Stichproben, z. B. aus dem HRS-Datensatz (n=13,254), benötigte die traditionelle Methode signifikant mehr Zeit. Allerdings war bei Verwendung der Scatterplot-Methode nur eine moderate zusätzliche Zeit erforderlich.

Die Scatterplot-Methode war auch in der Lage, CNV-Genotypisierungen auf dem X-Chromosom durchzuführen.

Zusammenfassend lässt sich sagen, dass die Scatterplot-Methode eine vielversprechende robuste Genotypisierung auf vordefinierte CNVs in großen Kohorten mit hoher Genauigkeit und Effizienz ermöglichte.