

Zusammenfassung

Krebserkrankungen gehören in Deutschland und weltweit zu den häufigsten Todesursachen. Unter Krebs versteht man eine Reihe von Erkrankungen, bei denen sich Zellen aufgrund von genetischen Mutationen ungehemmt vermehren. Mit zunehmenden Möglichkeiten in der Genomsequenzierung und Datenverarbeitung richtet sich der Fokus in der Krebsforschung auf die molekularbiologischen Prozesse.

Ein solcher Prozess ist das alternative Splicing. Dabei entstehen durch das Weglassen verschiedener Teilabschnitte eines Gens unterschiedliche Transkripte. Erst seit wenigen Jahren gibt es große Datensätze zu alternativem Splicing in verschiedenen Krebskohorten.

In dieser Arbeit wurde untersucht, ob es einen Zusammenhang zwischen alternativen Splice-Ereignissen, genauer gesagt Exon-Skipping-Events, und Brustkrebs gibt. Dafür lag ein Datensatz mit 1036 Beobachtungen und über 200.000 Exon-Skipping-Events (Variablen) vor.

Mit dem UMAP-Verfahren zur Dimensionsreduktion und Clustering-Verfahren wurde analysiert, ob man anhand dieser Daten bekannte Brustkrebssubtypen identifizieren oder neue Subtypen definieren kann.

Auf diese Weise konnte in der Brustkrebskohorte des Cancer Genome Atlas ein bisher nicht annotierter Subtyp entdeckt werden, der vermutlich neuroendokrinen Ursprungs ist. Dieser Subtyp unterscheidet sich nur in wenigen Exon-Skipping-Events von den restlichen Brustkrebstumoren. Weitere Analysen ergaben, dass anhand dieser Exon-Skipping-Events auch neuroendokrine Tumore in anderen Körperregionen identifiziert werden können.

Schlagwörter: UMAP, Clusteranalyse, alternatives Splicing, Brustkrebs

Abstract

Cancer is one of the most frequent causes of death in Germany and worldwide. The term cancer describes a group of diseases which are characterized by unregulated cell growth. With increasing possibilities in genome sequencing and data processing, cancer research focuses on molecular biological processes.

One such process is alternative splicing. During this process a single gene can create different transcripts by excluding different segments of a gene. Only since a few years, large data sets on alternative splicing in different cancer cohorts have been available.

In this work, it was investigated whether there is a relation between alternative splicing events, more precisely exon skipping events, and breast cancer. For this purpose a dataset with 1036 observations and more than 200,000 exon skipping events (variables) was available.

Using the UMAP method for dimension reduction and clustering methods, it was analyzed whether known breast cancer subtypes can be identified or new subtypes defined.

Using this methods, a so far unannotated subtype in the breast cancer cohort of the Cancer Genome Atlas was discovered, which is presumably of neuroendocrine origin. This subtype differs from the other breast cancer tumors only in a few exon skipping events. Further analyses showed that these exon-skipping events can also be used to identify neuroendocrine tumors in other body tissues.

Keywords: UMAP, Cluster analysis, alternative splicing, breast cancer