

Google Goes Cancer

Berechnete Wichtigkeit

Prof. Dr. med. Georg Hoffmann

Stopp! Bitte blättern Sie nicht gleich weiter, nur weil auf dieser Seite eine Formel steht! Diese Gleichung hat die Welt verändert.

$$PR_i = \frac{1-d}{n} + d \sum_{\forall j \in \{(j,i)\}} \frac{PR_j}{c_j}$$

PR bedeutet „PageRank“ und ist eine Maßzahl für die Wichtigkeit von Webseiten^[1]. Durch ihre Berechnung kam 1998 Ordnung in die – damals noch lächerlich wenigen – 25 Millionen Webseiten, die Google bei seiner Unternehmensgründung gelistet hatte. 2008 wurde die Grenze von einer Billion Seiten überschritten. Seither zählt niemand mehr mit.

Die Grundidee der Formel ist allen Lesern von Fachliteratur wohlvertraut, denn ganz ähnlich bestimmt man auch die Wichtigkeit von Publikationen: Je öfter eine Arbeit von anderen Autoren zitiert wird, desto wichtiger muss sie wohl sein. Anstelle von Literaturzitaten zählte Google-Gründer Larry Page für seinen PageRank die Hyperlinks, die von einer Website auf eine andere verwiesen.

Dieses Zählen wird durch das Summenzeichen in der obigen Formel ausgedrückt: Die Wichtigkeit PR_i auf der linken Seite ergibt sich durch schlichtes Addieren der „Wichtigkeiten“ PR_j der Hyperlinks auf der rechten Seite. Alles klar?

Der Autor des Beitrags *Google Goes Cancer* hat netterweise auf Formeln verzichtet, aber Sie können davon ausgehen, dass seine *NetRank*-Formel so ähnlich wie die obige aussieht.

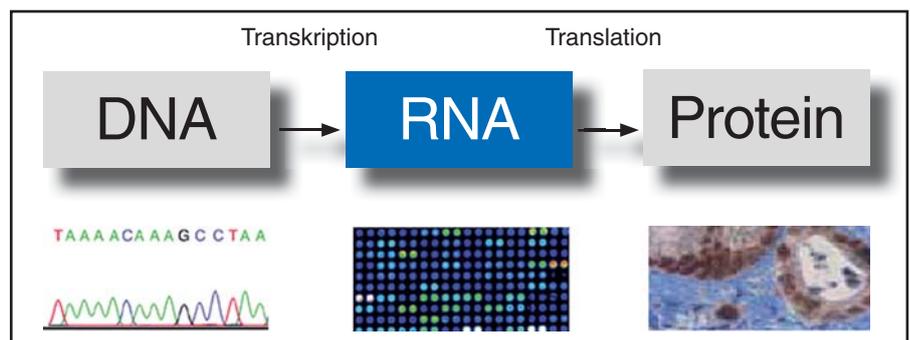
[1] Page L et al. The PageRank citation ranking: Bringing order to the web. Tech Rep Stanford Dig Lib Tech Proj 1998

Die erfolgreiche Suchmaschinenteknik von Google befruchtet neuerdings auch die Onkologie. Mit dem *NetRank*-Algorithmus lassen sich relevante Gene aufspüren, die prognostische Aussagen zur Überlebenszeit von Krebspatienten erlauben.

Von den rund 21.000 Genen des Menschen sind schätzungsweise 20 Prozent an der Weitergabe der korrekten genetischen Information von einer Zellgeneration zur nächsten beteiligt. Dieser gewaltige Aufwand wird benötigt, obwohl der Kopiervorgang der DNA-Reduplikation mit nur einem Promille Fehlerrate behaftet ist.

Molekularbiologisch können maligne Tumoren auf allen drei hier abgebildeten Ebenen charakterisiert werden, denn aus der Weitergabe fehlerhafter DNA (somatische Mutationen) resultiert eine gestörte Expression von RNA und Proteinen.

Seit über zehn Jahren werden insbesondere auf der Ebene der messenger-RNA



In der vorliegenden Studie wurden Tausende von mRNA-Konzentrationen mit Microarrays gemessen. Für die als relevant erkannten Biomarker wurden diese dann mittels PCR verifiziert und die Proteine zusätzlich immunhistochemisch nachgewiesen (Bildquelle: Trillium GmbH).

Bei insgesamt drei Milliarden Basenpaaren kommt man aber auf 30 Millionen Kopierfehler pro Zellteilung (!), und so gesehen erscheint es dann doch fast wie ein Wunder, dass unser Genom im Lauf des Lebens relativ konstant bleibt. Gäbe es keine Korrektur- und Reparaturmechanismen, so würden wir womöglich über die Jahre zu unvorhersehbaren Fabelwesen mutieren.

Durchaus Ähnliches geschieht bei der malignen Entartung: Hier geht die Kontrolle über die genomische Konstanz durch Mutation von „Wächtergenen“ verloren, so dass sich das Krebsgenom immer weiter von seinem Urzustand entfernt.

umfangreiche Genexpressionsstudien mit Microarrays durchgeführt. Man hofft, aus charakteristischen Mustern den weiteren Krankheitsverlauf vorhersagen zu können. In Einzelfällen wie dem Mammakarzinom ist dies gelungen, aber bei vielen hochaggressiven Krebsarten, etwa dem Pankreaskarzinom, besteht weiterhin hoher Forschungsbedarf.

Zwischen gesunden und malignen Zellen findet man meist erhebliche Differenzen im Genexpressionsmuster; dagegen unterscheiden sich günstige und ungünstige Verlaufsformen ein und derselben Krebsart oft nur geringfügig. Entsprechend dürftige

