

Bioinformatik –
Gespräch mit Joachim Selbig, Uni Potsdam

Rechnen gegen Krebs

■ Joachim Selbig ist Bioinformatiker an der Uni Potsdam und dem MPI für molekulare Pflanzenphysiologie in Potsdam-Golm. Für das deutsch-slowenische Projekt „SYSTHER“ analysiert er mit diversen „Omik“-Technologien erzeugte Daten. Ziel: Die Entstehung von Krebs zu verstehen und die Behandlung zu verbessern.

Laborjournal: Was wird in „SYSTHER“ untersucht?

Joachim Selbig: Wir erforschen systembiologisch die Entstehung und den Verlauf der häufigsten Gehirntumoren, die Gliome. Dafür wird nahezu die Gesamtheit aller molekularen Komponenten in gesunden und Krebszellen erfasst: Die Sequenzdaten – das Genom –, die Expression der Gene – das Transkriptom –, die Proteinkonzentration – das Proteom – sowie die Stoffwechselprodukte – das Metabolom. Besonders von der Metabolomik, an deren Entwicklung das MPI in Golm maßgeblich beteiligt ist, verspricht man sich neue Ansätze für Diagnostik und Therapie. Denn die Umwandlung von normalen Zellen in Tumorzellen geht mit starken Veränderungen des Stoffwechsels einher.

Wer ist an dem Projekt beteiligt?

Selbig: Das Nationale Institut für Biologie und das Bluttransfusionszentrum in Ljubljana, Slowenien, die Humboldt-Universität in Berlin, die Neurochirurgische Klinik in Großhadern bei München, die Universität Potsdam und die Berliner Firma MicroDiscovery. Während die in der Klinik tätigen Arbeitsgruppen für die Bereitstellung und Untersuchung des Patientenmaterials zuständig sind, besteht die Aufgabe meiner Gruppe und der Firma MicroDiscovery darin, die produzierten Daten zu verwalten und zu analysieren. Da wir das Zusammenspiel der einzelnen Zellbestandteile miteinander aufdecken wollen, analysieren wir die Ergebnisse nicht voneinander getrennt, sondern in ihrer Gesamtheit.

Wie machen Sie das?

Selbig: Wir bauen als Erstes eine Datenbank auf, die berücksichtigt, dass die Daten komplexer und heterogener Natur sind. Experimentelle Analysemethoden messen ja unterschiedlich genau und weisen daher voneinander abweichende Messfehler auf. Außerdem variieren die Konzentrationen der nachgewiesenen Moleküle in der Zelle extrem: Ein Gen verändert seine Konzentration in einer ganz anderen Größenordnung als eine Aminosäure oder ein Zucker. Wir müssen daher die aus den verschiedenen Mes-

„Wir erstellen mit Hilfe der [Next Generation und Deep Sequencing-]Daten mathematische Modelle, die uns dabei helfen sollen, die Entstehung von Krebs systembiologisch zu verstehen.“

sungen gewonnenen Daten mit Hilfe von statistischen Verfahren miteinander vergleichbar machen. Diesen Vorgang nennt man Normalisierung. Zum Teil stehen uns hierfür bereits statistische Methoden zur Verfügung. Zusätzlich entwickeln wir auch neue.

Sprengt eine solche Datenbank nicht schnell die Leistungsfähigkeit der Rechner?

Selbig: Glücklicherweise schreitet die Rechentechnik so rasch voran, dass bei der Hardware der Datenanalyse – der Speicherkapazität und Rechengeschwindigkeit – fast keine Flaschenhälse mehr bestehen. Allerdings gibt es hier eine Einschränkung: Die neuen als „Deep Sequencing“ oder

„Next Generation Sequencing“ bezeichneten Sequenziermethoden, die auch unsere Kooperationspartner in Großhadern nutzen werden. Mit Hilfe dieser Technologien können winzige Unterschiede zwischen Gensequenzen erfasst werden, die so genannten „Snips“ (SNPs – *Single Nucleotide Polymorphisms*), die für Diagnostik und Therapie wichtig sind. Diese Sequenziermethoden produzieren riesige Datenmengen, deren Verwaltung noch keine Routine ist. Sie stellt eine Herausforderung dar, die viel Software-Entwicklung erfordert.

Wie werden die Daten in die Datenbank eingegeben?

Selbig: Die an „SYSTHER“ beteiligten Partner können ihre Daten in die von uns in Kooperation mit MicroDiscovery entwickelte Datenbank eingeben. Beim Einleseprozess werden die Dateiformate der unterschiedlichen Messverfahren automatisch in ein Datenbank-kompatibles Format umgewandelt. In dieser neuen Form können die Daten später effizient weiterverarbeitet und ausgewertet werden.

Was kann man nach der Dateneingabe aus der Datenbank ablesen?

Selbig: Man kann sich anhand von verschiedenen graphischen Darstellungen – etwa zur Verteilung oder Häufung von Messwerten – einen ersten Eindruck über die Zusammenhänge zwischen Datensätzen verschaffen. Bei diesem ersten Blick sieht man allerdings „Wolken“ aus Daten auf dem Bildschirm. Denn diese Graphiken zeigen Tausende von Unterschieden an, etwa zwischen den insgesamt 25.000 menschlichen Genen gesunder und kranker Zellen. Die gleiche Zahl gilt auch für Abweichungen zwischen Proteinen oder Stoffwechselprodukten, den Metaboliten. Das menschliche Auge ist viel zu ungenau,

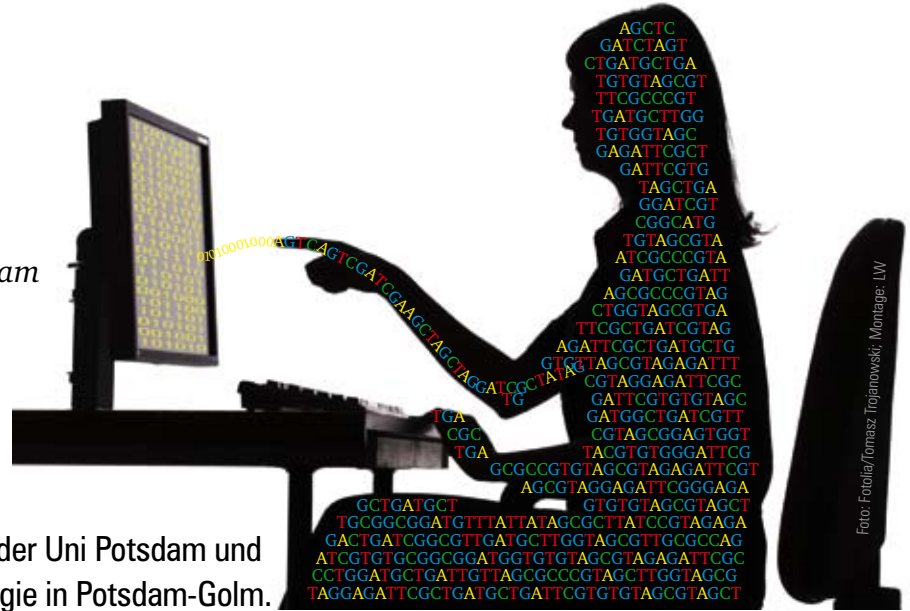


Foto: Fotolia/Tomasz Trojanowski; Montage: LW

um die Feinheiten einer solchen Darstellung zu erkennen. Um das Bild genauer aufzulösen, benötigen wir daher weitere Datenanalyseschritte, die außerhalb der Datenbank stattfinden: Sie filtern aus den vielen Punkten die wichtigsten heraus, die den Übergang von einer normalen Zelle zu einer Tumorzelle charakterisieren.

Wie funktioniert diese Auslese?

Selbig: Wir setzen hierfür mathematische Verfahren ein, beispielsweise die so genannte Dimensionsreduktion. Sie berechnet für die verschiedenen Zellmerkmale – Gene, Proteine, Metabolite – eine Rangfolge der Wichtigkeit bei der Tumorausbildung. Nach einer solchen Filterung der Daten bleiben von rund zweitausend unterschiedlich ausgeprägten Genen einer krankhaft veränderten Zelle im Vergleich zu einer gesunden Zelle nur noch 100 bis 200 Gene übrig. Auch die Werte für Proteom, Transkriptom und Metabolom lassen sich auf diese Weise verringern.

Was geschieht mit den gefilterten Daten?

Selbig: Wir erstellen mit Hilfe dieser Daten mathematische Modelle, die uns dabei helfen sollen, die Entstehung von Krebs systembiologisch zu verstehen: Welche Gene lösen die zellulären Veränderungen aus? Was bewirkt die Expression dieser Gene auf der Proteomebene? Wie beeinflusst die abweichende Proteinexpression schließlich den Umsatz der Metabolite, also den Stoffwechsel der Zelle? Wir beginnen hier erste Antworten auf alte Fragen zu finden.

Wie helfen die Daten bei der Optimierung der Diagnose und Therapie von Krebs?

Selbig: Wenn man aus dem Berg von Daten die wichtigsten Moleküle für die Transformation von gesunden zu Tumor-

zellen herausgefischt hat, können diese Moleküle als Biomarker genutzt werden. Mit Hilfe dieser Marker kann man Erkrankungen aufspüren. Darüber hinaus erlauben sie es, das Ansprechen auf Therapien zu beobachten. Bei einem Therapieversagen kann so die Behandlung zügig umgestellt werden. Ideale Biomarker lassen sich nicht nur im Gewebe, sondern auch im Blut von Patienten nachweisen. Denn dort sind sie gut zugänglich und können schon frühzeitig krankhafte Veränderungen anzeigen. Die Entwicklung neuer Krebsheilmittel ist ein weiteres Ziel. Der entgleiste Stoffwechsel der Tumorzellen bietet eine neue Angriffsfläche für Medikamente: Sie könnten beispielsweise die Versorgung der Tumorzellen unterbinden.

Wie lange läuft das SYSTHER-Projekt?

Selbig: Das Projekt wird seit 2007 vom Bundesministerium für Bildung und Forschung (BMBF) gefördert und wird noch zwei weitere Jahre laufen. Bis zum Abschluss des Projekts werden sich in der „SYSTHER“-Datenbank Datensätze aus Experimenten mit Zelllinien sowie Patientenmaterial befinden. Bisher schränken die Kosten, beispielsweise für die Ermittlung der Genexpression mit Hilfe von DNA-Microarrays, die Menge der untersuchten Proben ein.

Planen Sie die „SYSTHER“-Datenbank mit anderen Datenbanken zu vernetzen?

Selbig: Es gibt einen regen Austausch zwischen unserem und ähnlichen Projekten. Viele der Daten sind frei verfügbar, auch ein Teil unserer Daten. Daher sind wir stets bemüht, für unsere mathematischen Analysen statistische Methoden zu verwenden und zu entwickeln, die für jedermann frei zugänglich sind. Die Nutzung des Inhalts der „SYSTHER“-Datenbank in einer weiteren übergeordneten Datenbank

wäre wünschenswert. An dieser Möglichkeit wird derzeit auch bereits gearbeitet.

Gibt es schon Beispiele für Therapieverbesserungen durch systembiologische Ansätze?

Selbig: Am Saarbrücker Max-Planck-Institut für Informatik leitet Thomas Lengauer das Projekt „Geno2Pheno“, an dessen Initiierung wir mitbeteiligt waren. In diesem Projekt werden verschiedene Daten von HI-Viren und ihrem Wirt erfasst und miteinander verrechnet, um so die Wirksamkeit von antiviralen Medikamenten vorhersagen zu können und die Therapie von HIV-infizierten Patienten zu verbessern. Derartige individuell angepasste Therapien sind auch in anderen medizinischen Bereichen denkbar, etwa den Tumorerkrankungen. Bei der Entwicklung völlig neuartiger Therapien ist man bislang noch in der Entwicklungsphase. Hier kann man noch nicht mit ersten Erfolgsergebnissen aufwarten.

Wie schätzen Sie die Fortschritte auf diesem Gebiet ein? Wird man komplexe Erkrankungen, wie Krebs, bald vollständig verstehen?

Selbig: Ich glaube, wir sind gut beraten, vorsichtig mit unseren Erwartungen zu sein: Wir finden durch unsere Vorgehensweise zwar viele Antworten auf lang gestellte Fragen. Aber mit jeder Frage, die wir beantworten können, taucht eine Vielzahl neuer Fragen auf. Folglich ist es schwer zu sagen, wann es zu einem wirklichen Verständnis der Grundlagen dieser Krankheiten kommen wird und wir diese vollständig heilen können. Mit Sicherheit jedoch werden sich die diagnostischen und therapeutischen Möglichkeiten dank der systembiologischen Ansätze verbessern.

INTERVIEW: MELANIE ESTRELLA