

Sebastian Schönherr, Günther Specht und Lukas Forer (v. li.) entwickeln mit den Partnern Lösungen zur Verarbeitung komplexer Datenmengen.

DER GENETISCHE CODE IN DER DATENWOLKE

Die sprichwörtliche Nadel im Heuhaufen sucht eine interdisziplinäre Kooperation zwischen der Universität Innsbruck um Prof. Günther Specht und der Medizinischen Universität um Prof. Florian Kronenberg im menschlichen Genom. Informatik und Genetik als Beispiel einer vorteilhaften Symbiose.

Seit das menschliche Erbgut vor über zehn Jahren erstmals entziffert wurde, haben sich die Technologien zur Sequenzierung von Genomen rasant weiterentwickelt. Während im Rahmen des ersten Humangenomprojekts noch über 1000 Forscherinnen und Forscher aus 40 Ländern jahrelang an der Entschlüsselung der menschlichen Gensequenz gearbeitet haben, lässt sich heute das gesamte Genom eines Menschen in wenigen Stunden maschinell sequenzieren. Schon ist die Rede davon, die Sequenz des Genoms über einen an die USB-Schnittstelle angeschlossenen Mini-Sequenzierer im Taschenformat in kürzester

Zeit aus einem Blutropfen auszulesen. Dabei sammeln sich gewaltige Datenmengen an. Denn jedes menschliche Genom besteht aus 3,3 Milliarden Basenpaaren, deren Zusammensetzung in einer Buchstabenfolge beschrieben wird. „Unsere Projektpartner an der Medizinischen Universität arbeiten hauptsächlich nur mit einem Prozent des tatsächlichen Erbguts und dennoch enthält der Datensatz eines Patienten bereits viele Gigabyte an Informationen“, erklärt Sebastian Schönherr aus der Forschungsgruppe Datenbanken und Informationssysteme um Prof. Günther Specht am Institut für Informatik. Die Datenbank-Spezialisten arbeiten dabei

eng mit der Arbeitsgruppe um Prof. Florian Kronenberg und Ass.-Prof. Anita Kloss-Brandstätter von der Sektion für Genetische Epidemiologie zusammen. Dort wird im großen Maßstab mit Genomen gearbeitet, gilt es doch durch den Vergleich des Erbguts sehr vieler Menschen mögliche Verbindungen zwischen genetischen Veränderungen und Krankheiten herauszufinden. So landen die anonymisierten Daten von tausenden Patienten und Gesunden auf den Rechnern der Innsbrucker Wissenschaftler. „Dies stellt uns vor enorme technische Probleme, an deren Lösung wir seit fünf Jahren gemeinsam arbeiten“, erläutert Günther Specht.

INTERNATIONALER STANDARD

Das Innsbrucker Forscherteam, in dem auch die beiden Informatiker Hansi Weißensteiner und Lukas Forer mitarbeiten, entwickelte in einem ersten Projekt eine Datenbank zur kombinierten Speicherung von Phänotypen (dies ist das „Erscheinungsbild“ eines Menschen im Sinne von Erkrankungen und anderen Eigenschaften wie Labormesswerten) und der genetischen Ausstattung. Dies erlaubt die gleichzeitige Analyse von Gendefekten und den daraus resultierenden Krankheiten. Dazu werden die anonymisierten Daten von Patienten in die Datenbank eingespeist und analysiert. Speichern und Rechnen sind die beiden großen Herausforderungen bei der Verarbeitung der Genomdaten. Dabei muss sehr gut überlegt werden, wie die Informationen abgespeichert werden. Denn ohne einen effizienten Zugriff verlängert sich die Laufzeit von Berechnungen dramatisch. Für die Analyse der Daten sind darüber hinaus effiziente Programme notwendig. „Wir sind hier auf massive Parallelität angewiesen“, betont Günther Specht. „Diese enormen Datenmengen können nur in einer Rechnerwolke – einer sogenannten Cloud – sinnvoll verarbeitet werden.“ Dazu steht am Institut für Informatik ein Rechner-Cluster im Sinne einer riesigen Anzahl von verbundenen Einzelrechnern zur Verfügung. Durch eine Forschungsförderung hat sich den Informatikern zusätzlich die Chance geboten, ihre Lösungen auch auf kommerziellen Cloud-Anbietern wie Amazon auszuführen, die den Zugang zu verteilten Rechnersystemen für Wissenschaftler weltweit erleichtern. Das Ziel des Innsbrucker Forscherteams ist es, eine Plattform zu erschaffen, die es Wissenschaftlern erlaubt, ihre eigenen Programme auf einem Cluster auszuführen, den Transfer von riesigen Datenmengen zu ermöglichen und die Reproduzierbarkeit von Ergebnissen zu gewährleisten.

Besonderes Augenmerk stellt hierbei die einfache Bedienbarkeit der Plattform und somit die Verwendbarkeit der inte-


grierten Programme für Wissenschaftler dar. „Das ist entscheidend, um eine gute Akzeptanz bei den Nutzerinnen und Nutzern im Bereich der Genetik zu erreichen“, unterstreicht Sebastian Schönherr die Notwendigkeit dieses Projektes.

„Wir stellen somit eine Plattform frei zugänglich zur Verfügung, die auch von anderen Wissenschaftlern verwendet werden kann“, so Günther Specht. Aus aller Welt greifen Forscherinnen und Forscher regelmäßig auf zahlreiche der Innsbrucker Open-Source-Produkte zu, in denen bereits Anwendungen mehrerer Forschungsgruppen integriert wurden.

Mit ihrer Software versuchen die Forscher der beiden Innsbrucker Universitäten in diesem rasant wachsenden Forschungsfeld einen Standard zu etablieren. „Die Reproduzierbarkeit von Ergebnissen war früher nicht immer gewährleistet oder nur mit sehr großem Aufwand möglich“, sagt Prof. Specht.

WO KOMMEN WIR HER?

Die genetischen Informationen des Menschen geben aber nicht nur Aufschluss über mögliche Krankheitsrisiken, sondern lassen auch Rückschlüsse auf die Evolution der Menschheit zu. Kommen unsere Vorfahren wirklich aus Afrika? Und wie haben sich die Menschen auf der Welt ausgebreitet? Um diese Fragen zu klären, wird die Erbinformation aus den Energiekraftwerken der menschlichen Zelle – den Mitochondrien – entziffert. Dieses Erbgut wird nur über die Mutter an den Nachwuchs weitergegeben und erlaubt daher die Rückverfolgung der Menschheit auf ihre Ursprünge vor tausenden von Jahren.

Auch für dieses Problem haben die Informatiker um Günther Specht in Zusammenarbeit mit Anita Kloss-Brandstätter eine Software entwickelt, die eine automatische Zuordnung von Personen zu sogenannten Haplogruppen erlaubt. Diese Gruppen geben Auskunft über das verwandtschaftliche Verhältnis von Menschen. Die Arbeit wird von Forschern weltweit verwendet und erleichtert das Bestimmen der Haplogruppe von der mitochondrialen DNA um ein Vielfaches. „Bisher verglich man das mitochondriale Erbgut mit einer Referenzperson manuell“, erzählt Hansi Weißensteiner. „Nun haben wir diese aufwendige Feinarbeit automatisiert, was von der internationalen Forschergemeinde sehr gut angenommen wird.“ So findet die Software starken Anklang im Bereich der klinischen Genetik, um Krankheitsrisiken von Mutationen festzustellen, in der Populationsgenetik, die versucht Entwicklungsmuster des Menschen zu entdecken sowie in der Gerichtsmedizin zur Identifikation von Personen. cf 

FACTS



Seit 2007 entwickeln Forscher der Innsbrucker Universitäten gemeinsam Werkzeuge zur Analyse von genetischen Informationen. Dabei arbeiten Informatiker, Statistiker, Biologen und Mediziner eng zusammen. Sie profitieren gegenseitig von der fachübergreifenden Lösung von Problemen. Die entwickelte Software spart viel Zeit bei der Arbeit mit genetischen Daten und macht Analysen möglich, die bisher undenkbar waren. Für die Zukunft gilt es, die Konzepte und Programme an die sich rasant weiterentwickelnden Technologien der Genomforschung ständig anzupassen.