

# Big Data in der Genetik

Innsbrucker Informatiker und Genetiker entwickelten eine Plattform, mit der große Mengen an genetischen Daten in kurzer Zeit verarbeitet werden können.

Das „next generation sequencing“ brachte die weltweite Forschergemeinde zwar schneller an das Objekt ihrer Begierde, die neuartigen DNA-Sequenziergeräte stellten die Wissenschaftler aber auch vor ein Problem: Was tun mit genetischer Big Data?

„Als wir uns erstmals mit der Verwaltung von biologischen und genetischen Daten beschäftigten, war von Big Data noch keine Rede“, erinnern sich Sebastian Schönherr und Hansi Weißensteiner. Die zwei Informatiker suchten 2006 ein Diplomarbeitsthema: „Günther Specht, Professor für Datenbanken und Informationssysteme an der Uni Innsbruck, hat uns auf ein Projekt an der Medizinischen Universität hingewiesen.“ Dort trafen sie auf Florian Kronenberg. „Wir verstanden ihre Sprache nicht, sie nicht die unsere. Wir mussten erst lernen, miteinander zu kommunizieren“, lacht der Leiter der Division für Genetische Epidemiologie. Die Kommunikation wurde perfekt und mündete nicht nur in Diplomarbeiten und Dissertationen, sondern in die Open-Source-Plattform Cloudgene, auf die Forscher der ganzen Welt zugreifen.

2007 hatten Schönherr und Weißensteiner ein Informationssystem aufgebaut, in dem genetische und phänotypische Daten gespeichert wer-



Foto: Andreas Friedle

Genetiker Florian Kronenberg (2. v. re.) mit seiner Informatik-Abteilung, die hinter Cloudgene steckt: Sebastian Schönherr; Hansi Weißensteiner und Lukas Forer (v.li.).

den konnten. 2009 stellte sich bei der Diplomarbeit von Lukas Forer – ebenfalls Informatiker – die Frage, „wie man die Datenmengen effizient parallel analysieren und speichern kann.“ Die Lösung fanden sie in einem – zuerst aus alten Geräten zusammengesetzten – Computer-Cluster, mit dem aber, so Schönherr, „zu kommunizieren für einen Nicht-Informatiker mühsam ist“. „An der Schnittstelle zwischen Parallelität und Usability entstand Cloudgene“, ergänzt Forer. Der User dockt am Cloudgene-Cluster an, lädt seine Daten, z.B. die sequenzierte DNA von 2000 Probanden, ins

System und kann diese am normalen PC-Arbeitsplatz analysieren und parallel komplizierte Berechnungen laufen lassen. Verschiedene Anwendungsfälle sind in der Zwischenzeit in die Plattform integriert und werden als Service kostenlos zur Verfügung gestellt. Eine davon ermöglicht mit Hilfe von mitochondrialer DNA eine automatische Zuordnung von Personen zu sogenannten Haplogruppen, womit verwandtschaftliche Verhältnisse von Menschen festgestellt werden können. Eine Frage, die offensichtlich interessiert: „Monatlich haben wir rund 1500 Zugriffe – und das weltweit.“ ]