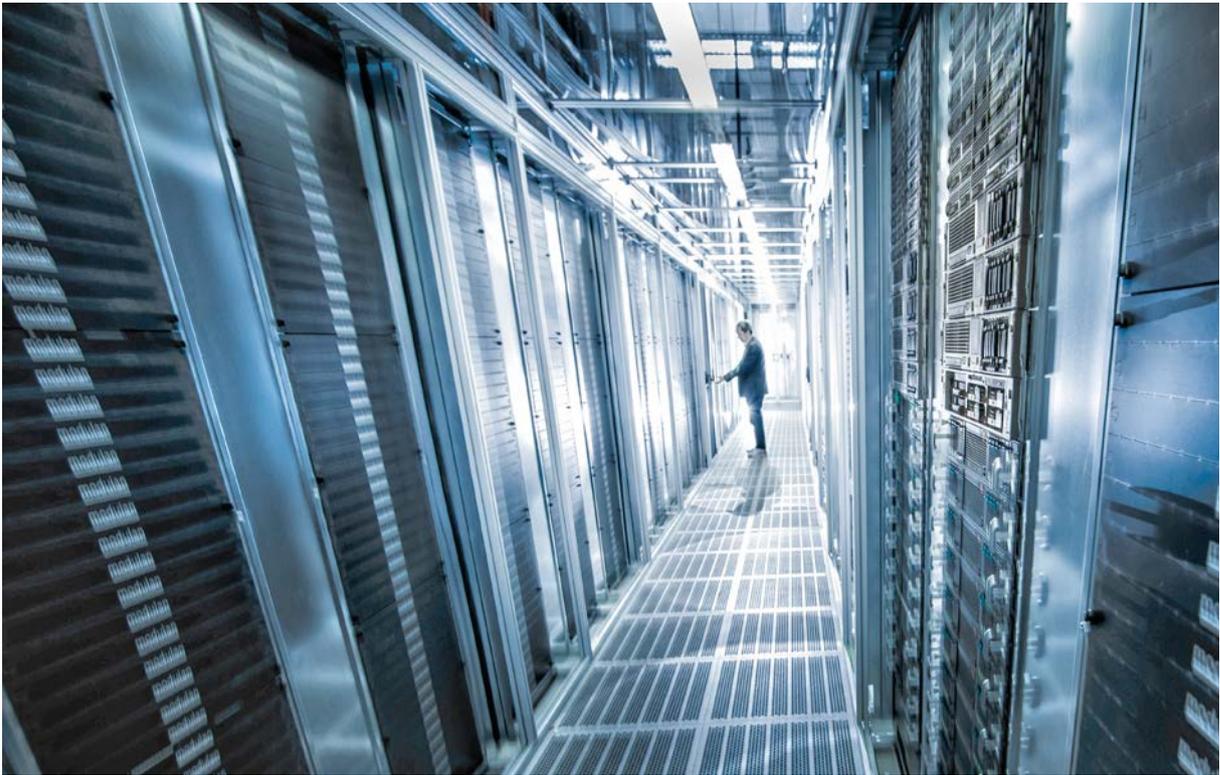BIOINFORMATICS FOR BETTER THERAPIES

Data clouds in medicine

IT centers for huge data volumes: supercomputers help scientists analyze medical data. Their use can reveal previously unsuspected interrelationships.

The mountains of data generated by research and modern medicine contain valuable information, such as previously unrecognized correlations between different mutations or other biomarkers and the therapeutic success of a drug product. A certain combination of these gene variants – or alleles, as they are known – could increase the patient's likelihood of responding well to a cancer drug. "If these predispositions were known, doctors could precisely select the right treatment for the patient in question," says Dr. Joerg Lippert, head of Clinical Pharmacometrics in Bayer's Pharmaceuticals Division. However, this information must first be extracted from the data. "We have to use all of the information in order to be able to make the best decisions possible in modern medicine," explains Lippert.

The amount of medical data currently generated by routine diagnostics and medical studies is already almost unmanageable. "Big data has long been a reality in medicine. The three Vs – volume, velocity and variety – will increasingly determine everyday reality in doctors' offices," says Lippert. Volume refers to the quantity of data and velocity to the speed with which it is generated. Variety, or complexity, is a particularly challenging factor in medical data. For example, a patient's record contains not only measurements and tables but also diagnostic images –

the results of a wide range of examination methods. "At present, medical data is typically still very unstructured and in some cases inaccurate."

Scientists first have to make unstructured medical data usable

"We therefore have to convert the data into a form that can be processed by a computer manually or with the assistance of special computer algorithms," explains Lippert. Once conversion is complete, the researchers rely on their algorithms and computers. "We let the data speak for themselves, meaning we start off with as few assumptions as possible. This prevents our expectations from limiting our analyses, and the results are nearly entirely open. This can lead us to new hypotheses," says Lippert. Due to the large quantity of data, this approach requires high processing power. "If we have a data set with 50,000 patients containing 5,000 health parameters for each of them, that results in an astronomical number of combinations to be examined," explains the Bayer scientist. This is why the researchers are relying on heuristics – special methods that help them structure data without requiring too many assumptions in advance. It is

a compromise solution that enables complex computing operations to be performed in a reasonable amount of time. "We are searching for statistical correlations in huge sets of data. That necessitates machine learning, an approach we have been pursuing for years now. The main difference is that today the data sets are larger and the computers are faster, which leads to a new level of quality," explains Lippert.

In fact, the data specialists can already today estimate the ideal doses of drug products using computer tools, an innovation that is particularly valuable when it comes to planning clinical trials. "With the right data, it can save us several years of development time. That helps the patients, because we can get a new therapeutic option to them faster," sums up Lippert. The new approaches that he and his team enabled with their work made it possible, for instance, to omit a specific study section of Phase II clinical development for a drug to treat heart failure. They were therefore able to save more than a year of development time. The data experts are very much still at the beginning but Lippert is nonetheless convinced: "We can help shape the future of medicine and ensure better therapies." ∎

Numbers, formulas and models: Dr. Joerg Lippert's team is searching through huge medical data sets to find new correlations. His findings could help doctors make better decisions.

"Big data: major legal challenges"

research *spoke with Sigrid Achenbach, Senior Counsel Law of the Pharmaceuticals Division at Bayer, about the legal and socio-political challenges associated with big data analyses in medicine.*

Sigrid Achenbach

What legal questions need to be clarified?

Analyses of disease-related patient information are particularly interesting for companies developing new drug products. However, they also collide with three of the principles of data protection legislation. Firstly, data processing must be fundamentally permitted – the principle of lawfulness – and should comprise as little personal data as possible – the principle of data minimization. In addition, this personal information may only be used for the agreed study – the principle of purpose limitation. In practice, it is very difficult to reconcile all these conditions in big data studies.

Quite a challenge! How can we solve it?

There is no universal solution at present, and it generally comes down to complicated case-by-case decisions. Even the EU's new General Data Protection Regulation that will enter into force in May 2018 is unlikely to resolve these difficulties. One potential solution would be broader patient consent permitting big data analyses. However, there are legal limits to this. Another possibil-

ity would be an independent data protection watchdog that could review and approve research projects going beyond the original agreement, but at the moment that is just a vision. This matter can only be resolved by all interest groups working together. That includes the responsible ministries and pharmaceutical companies, but also academic institutions and patient organizations.

How will this particular field of medicine develop in the next 20 years from now?

Research will increasingly make use of data from different sources and perform complex analyses. This will lead to new findings that help both patients and society as a whole. However, security measures will be in place to prevent misuse of sensitive data: thus, preventing individual patients from being identified and medical information from falling into the hands of third parties. It is important to me personally that only data supplied voluntarily are used – everyone should be able to decide for themselves what happens with their personal information.