Screening enormous databases to find a cure for cancer

March 9 2016

Developing new drugs is becoming increasingly difficult: research and lab experiments are time consuming and expensive. By making use of data first – screening large databases in order to discover patterns - we can make better predictions about what will work and what will not. This is what Gerard van Westen calls Data-Driven Drug Discovery. He received an NWO Veni grant for his current research, in which he uses data mining techniques to find a cure for cancer.

What exactly is your research focused on?

'I study G-Protein Coupled Receptors (GPCRs), which are proteins inside the cell membrane. GPCRs can be used to manipulate the cell from the outside, which is very useful for medications, for instance. Many drugs, such as antidepressants and antihistamines, work through GPCRs. My research is focused on whether GPCRs can be used as drug targets for the treatment of cancer. So far, research on anti-cancer drugs has mainly been focused on the use of a different type of proteins: the kinases. That makes sense, since kinases manage the growth of the cell. But I believe investigating the role of GPCRs may also have very useful results.'

What makes your research data-driven?

'In recent years, more and more large public databases have become available, in which researchers and pharmaceutical companies share their [data](link). An example is the ChEMBL database, which contains large amounts of data on experimental molecules in [drug discovery research](link). I analyze that data, using algorithms, in order to discover patterns. For instance, I try to find out how these experimental molecules or drug bind to proteins. Based on that, I can make predictions; for example, that a particular medicine may work very well with a particular protein. The next step is taking it to the lab in order to test the hypothesis. And if the results are good, a pharmaceutical company can start developing a medicine.'

So by making use of data we can better target our research?

'Exactly. If we want to develop good medicines, we should use what we already know as a starting point. And by that I do not mean reading three

or four articles, but really consulting the databases and using [data science](#) to do large-scale analyses. By doing that, we can generate much better hypotheses. Another advantage of data-driven research is that it leaves less room for the researcher's gut feeling. Everyone has personal preferences and is more familiar with one technique over the other. That is quite normal, but it often makes us think in a certain direction. By looking at predictions made by computers, we may see connections that we might never have thought of otherwise.'

What are the drawbacks of using such large amounts of data?

'One of the risks of data-driven research is that we start focusing too much on data and models, while losing sight of the goal: drug development. Analysing all this data is nice, but in the end we do it to make medicines. And then of course there are privacy issues: we can work with genetic information, which tells more about people than they might be comfortable with. Of course, there are strict rules for the use of this type of data, so I believe there is little risk that something goes wrong - at least in my field. Yet, many people may be reluctant to provide that kind of data. Unfortunately so, because the more data we have, the better us researchers can do our jobs.'

How has data science changed your area of research in recent years?

'The developments are moving fast. Computing power is growing rapidly: the same data set that took me a full week to compute during my internship, I was able to compute in only a few hours four years later. Also, algorithms are getting better and better. The latest development is deep learning: a technique in which we work with smart algorithms that behave in way similar to the human brain. This allows us to process data

much more efficiently, so we can better and faster predict interactions between proteins and drugs. The pharmaceutical industry has also realized the immense possibilities in the field of data research, and is increasingly seeking cooperation with academia.'

How do you see the future of Data-Driven Drug Discovery?

'The beauty of drug research is that it is a finite space. There are lots of molecules that could be drugs, but ultimately, their amount is limited. That means we could, theoretically, have a computer look at all these molecules, and then make all possible medicines based on the computer's findings. We are still a long way from having enough data to accomplish that, but perhaps sometime in the future.'

What would we need to do to achieve that?

'We should adopt new ways of dealing with data. Researchers should save all the data they generate and make it available to others. Now, a lot of data is thrown away, especially if a project has disappointing results. That is a shame, because it may be very useful to other researchers. By sharing data, we can build on each other's research more efficiently. A change is already going on: more and more researchers include data and algorithms as supplements to their publications. And also pharmaceutical companies share increasing amounts of data, for example in the projects Open PHACTS. But there is still much to be gained in this area.'

Provided by Leiden University