Finalist for OBR science blog competition 2014

A Hitchhiker's Guide to Genalice's Wonderland

By Dr. Quido Valent – QV Strategic Support

Once upon a time, not even so very long ago, this would have been a story where science fiction meets fairy-tale. Bert, first raised but later excommunicated by the wise and ancient breed of Sea-Mensch, was wandering the galaxy in search of a new life with only a towel for luggage and an inherited fascination for the number 42. He teamed up with wizard Hans, who was sent by the Oracle to help Bert on his quest. Their adventures would lead to exotic places, involve a prophecy, a challenge, the wizard's hat, wizard Hans becoming *The Wizard of Jos* (and many others), and of course the number 42. But this is no sci-fi fairy-tale. This is a tale of fairly new science that is bringing the promises of 'personalised medicine' closer to reality.

Wizard Hans' hat at work
(images with courtesy of Genalice)

'Big data' and 'personalised medicine' are *the* buzz-words promising custom fitted healthcare, the key opinion leaders in biomedical research predict. Scientists and clinicians subtract an enormous amount of data from individuals, including many 'omics', such as genomics (complete genetic profiles). Omics have the potential to ultimately move healthcare towards prevention or early treatment[1]. In addition, clinicians collect anamneses, perform diagnostic tests, order X-rays or fancy scans, but also data from psychoanalysis, dietary analysis, and much more. The enormous stream of information can in theory be used to uncover trends and abnormalities enabling prediction of an individual's predisposition for certain diseases, disease course, and/or response to treatment. Such insights will also give rise to new avenues for drug development.

But there is a catch. We lack the computer power and time to convert the continuous stream of data in many (often incompatible) formats into comprehensive and useful information. It took scientists many years and about one hundred million dollar to sequence the first human genome. Nowadays, with the newest Next Generation Sequencing tool, the HiSeq X Ten Sequencing System, Illumina promises to be able to generate the raw sequence data for only one thousand dollar per genome, at a rate of 18,000 sequences a year[2]. Splendid! But this data needs to be processed and analyzed further to generate useful information. Currently that takes several days per sequence. So, with one set of machines producing 18,000 raw sequences in 365 days – you do the math. This example is illustrative for many of the hurdles preventing the optimal use of big data.

The Dutch start-up Genalice has defined its very ambitious mission: *To save lives and increase the quality of life of people suffering from complex DNA diseases, such as cancer*. The founders, Hans Karten (CEO) and Bert Reijmerink (CSO) joined forces after leaving Oracle and Siemens, respectively. They stumbled on the big data issue illustrated above. The three most relevant hurdles were identified and the route to solving each was laid out. Hans and Bert did not want to improve what is already there. As Genalice they started paving a new path as they embarked on a quest to start from scratch, exploiting the currently available knowledge to the fullest.

Genalice's vision is to truly master all biomedical big data issues and provide software solutions that significantly accelerate i) preprocessing of Next Generation Sequencing data, providing individual genomic sequences, calling its variants and annotating these (tool: **MAP**), ii) integration and correlation of a large collection of human genomes and/or exomes with any other type of relevant (biomedical) data, thus significantly enhancing the discovery of disease causes and ultimately drug development (tool: **LINK**), and iii) the connection of all biomedical knowledge globally by allowing full

integration of this data into a Clinical Decision Support tool (**CHECK**) that will enable clinicians to select the most efficient treatment for each individual patient. In short: enabling the delivery of truly personalized medicine.

The company grew from a team of 2 to 14 persons in just two years. My guess is their end-goal is exactly 42 persons. Meanwhile, a third person has joined the Genalice management team. Jos Lunenberg (CBO) brings in the essential knowledge of the life sciences industry. A big part of his job is showing the world what the new technology can do. Jos understood that words are not enough. Thus, the team decided to take on the challenge to do the impossible.

State-of-the-art high-throughput technology, known as Next Generation Sequencing or NGS produces an enormous collection (hundreds of millions) of partially overlapping short sequences – the raw data. As a huge puzzle these pieces have to be properly aligned to ultimately yield the full genome sequence. Short-read alignment with the currently widely used Burrows-Wheeler Aligner (BWA) yields a single full genome sequence from raw NGS data in several days. Now the **challenge** was to use the new NGS short-read aligner GENALICE MAP and just one regular computer to process a staggering **42 human genomes[3] in a single day**. Yeah, right.

But they did it, hands down! Using just one dual Intel Xeon E5-2620 server they were able to map 42 full human genomes in just over 17,5 hours. With so much time left, they went on to align another 42 tomato genomes. The 84 genomes were mapped with 15 minutes left on the clock.

Impressive, but how? MAP uses an entirely new algorithm, designed by Hans on a Turkish beach. Part of the magic is related to the fact that he started from scratch and made optimal use of the modern hardware architecture. The details of this achievement, and a white paper on MAP can be found on the Genalice website. A video impression of project 42 was posted on YouTube.

As you understand, this was only step one in the quest to the Wonderland these visionary people want to create. Already strides are being made on the development of LINK as a beta version. What's more, the company aims to launch its first LINK release still in 2014. The final leap in sight is the development of the Clinical Decision Support software solution CHECK to support treating physicians in their daily clinical decision-making process. The team will focus on this from 2015 onwards.

Still, a healthy critical view on what is presented is advisable. Currently a solid, independently generated scientific foundation of all claims is lacking. Although the compatibility of MAP with existing tools is a big plus in light of quick implementation, before general acceptance can be expected an elaborate verification of particularly the claims on MAP speed, accuracy and reliability is required. Such information will come available in the near future as several academic and private parties are now testing MAP in real-life settings and more are likely to follow suit[4]. To this end, the company is already teaming up with top institutions like NIHR Oxford Biomedical Research Centre (UK) and Mount Sinai Hospital (USA). With all that promise, I know I will be following the team's progress closely as I am looking forward to living in Wonderland.

References:

1. EU COMMISSION STAFF WORKING DOCUMENT – Use of '-omics' technologies in the development of personalised medicine, Brussels, 25.10.2013, SWD(2013) 436 final
2. Illumina website
3. Raw sequences (FASTQ files) from different sources (for details hover your mouse over the 42 puppets on top of the webpage: http://42.genalice.com/event/)
4. Discover Magazine blog by Lauren Gravitz – Personalized Medicine Slogs Toward Reality, January 23, 2014