Measuring life: sensors and analytics for precision medicine

Stefan Harrer* 1, 2

1 IBM Research - Australia, 204 Lygon Street, 3053 Carlton, VIC, AU
2 University of Melbourne, 3010 Parkville, VIC, AU

ABSTRACT

The first industrial revolution focused on machines, the second one was data-centric – a third revolution combining the power of devices and information has just started and transforms our understanding of life itself. Thereby novel sensors and networks from wearable biometric devices to lab-on-a-chip platforms for exploratory fundamental research on single-biomolecule characterization and design occupy a key role. In combination with recent advances in big data analytics for life sciences, healthcare and genomics such sensors are essential tools for moving from fast and cheap personalized DNA-sequencing via smart genomics towards one-off prevention and treatment plans. Replacing state-of-the-art, one-fits-all approaches, this paradigm shifting individual “assess & response” scheme commonly referred to as precision medicine merges biomedical engineering, systems biology, systems genomics, and information technology. Integrated sensors for isolating, investigating and eventually manipulating single biomolecules are important experimental tools for developing next-generation DNA-sequencing platforms and for conducting ‘omics research which is a defining part of systems biology. In that context resistive pulse sensing has emerged as a powerful technology at the intersection of biotechnology and nanotechnology allowing electrical, label-free screening of biological compounds such as proteins or DNA with single-molecule, single-nucleotide and even single binding site resolution. Resistive pulse sensing technology has been at the center of recent commercial $100Ms investments in the next-generation DNA-sequencing sector. While next-generation sequencing platforms based on resistive pulse sensing techniques will mature further, the technology is also increasingly used for screening other biomolecules such as for example proteins. This allows for developing novel diagnostics and ultra-high throughput pre-clinical drug screening systems which might help to transform the pharma pipeline similarly to how the $1000-genome has revolutionized DNA-sequencing.

Keywords: resistive pulse sensing, precision medicine, systems biology, systems genomics, single molecule sensing, label-free, DNA-sequencing, drug screening, protein screening, nanopore, nanochannel

1. INTRODUCTION

A conventional engine control unit for commercial cars gathers dozens of data points per second. State-of-the-art flight data recorders collect around 3000 data points per second storing several gigabytes of data per flight. The so called APGAR test was developed in 1952 and to-date still is the most comprehensive check of a newborn’s organism: it is performed once in a baby’s life and collects a total of 5 data points. Why is it that we are monitoring and controlling systems which we entrust our lives to with such diligence, but do not do the same with our life itself? The answer lies in the complexity of the human organism: operating a car engine or flying an airplane is trivial compared to grasping the complexity of physiological interactions in the human body which shape every second of our life. Ever since the double-helix structure of DNA was discovered in 1953 we have investigated its role as the carrier of the code of life. It was a multi-$B effort over more than 10 years until a complete human genome had been fully sequenced for the first time in 2003. Being able to read the manual of human life was of course a breakthrough but information could not be derived from it. In order to make sense out of sequencing data it has to be individually contextualized with biomedical metadata (such as for example healthcare and medical records, medical family history, scientific publications). Being able to do this in a meaningful way requires three things: firstly, a statistically relevant number of reference cases needs to be compared to narrow down the physiological function of a specific gene, therefore secondly, DNA-sequencing needs to be fast and affordable for everyone, everywhere, and thirdly, given the size and inhomogeneity of the data to be correlated in a reasonable time, novel big data analytics technologies are needed.

*presenting speaker and author email: sharrer@au.ibm.com;
2. NOVEL SENSORS FOR PRECISION MEDICINE APPLICATIONS

Advancing DNA-sequencing from the $1000-Genome to the $1-Genome is a bionanotechnology hardware challenge residing in the field of biomedical device engineering. Setting up large-scale human genome studies to create and analyze sequencing and contextual medical data of an ever increasing number of individuals is a complex joint operational effort in the hands of patients, medical practitioners, scientists as well as regulatory and funding bodies. Developing platforms to process and analyze big data packages of unprecedented size and diversity is a software challenge calling for novel cognitive analytics techniques. The last 3 years have seen substantial advancements in all these areas: the size/cost footprint of DNA-sequencing tools is decreasing fast bringing cheap and portable DNA-sequencers into reach\(^1,2\). At the same time performance of high-throughput state-of-the-art sequencing techniques has improved at a pace that led to the abandonment of the X-Price for demonstrating the $1000-Genome as various teams are currently approaching and likely blazing through that milestone with the $100-Genome in sight. Building on these enhanced sequencing capabilities multiple private, commercial and government institutions have initiated large scale genome studies: ranging from a recent landmark study performed on the population of Iceland\(^3\) via the Genomics England Initiative\(^4\) and Craig Venter’s newly founded Human Longevity\(^5\) enterprise to the Precision Medicine Initiative of the US government\(^6\), current studies aim to collect whole genome information on patient pools consisting of several thousands to 100,000 genomes. At the same time advanced cognitive big data analytics tools like for example IBM’s Watson technology have emerged and mature showing first indications of early-stage commercial practicability\(^7\). Finally, in order to be able to respond to the findings of personalized genome analysis we need to develop faster and cheaper technologies for discovering novel drugs, for improving existing drugs, and for their intelligent use in an increasingly complex and customized environment transforming the pharma pipeline much like the $1000-Genome has revolutionized DNA-sequencing. First cases of all building blocks working in sync towards successful individual treatments have been reported recently\(^6,8\), but at this point these examples constitute encouraging, yet isolated efforts rather to proof that the concept can be moved from hype to impact, than demonstrating population-wide availability of precision medicine services.

Precision medicine is a story about data. Figure 1 schematically projects key data sources and corresponding sensor types onto the ‘omics canvas of precision medicine. While some data is readily available or can be extracted using existing techniques, other data has to be generated individually and from scratch requiring novel and improved sensing platforms. Besides biometric data creation through wearable sensors, this is naturally the case for sequencing data and also applies to data which we collect in the pre-clinical drug discovery phase to test the impact of drug compounds on physiological intra- and intercellular activities. It is our fragmentary understanding of the nature of these biomolecular cross functionalities that we mean when we talk about the complexity of the human organism. Investigating which proteins our body builds and why, how these proteins interact with each other and to which end, and finally how these interactions can be impacted by introducing drug compounds to the cellular environment is key to understanding, monitoring and eventually controlling our body with the same level of detail that we control a driving car or flying plane with. This task forms the backbone of systems genomics, systems biology and in particular ‘omics research. It covers the entire spectrum of the biomatrix from investigating biophysical characteristics and interaction dynamics of single molecules to analyzing the behavior of complex many-biomolecule systems in the context of the environment an individual lives in and the lifestyle it pursues. In that context resistive pulse sensing technology (often also referred to as nanopore/nanochannel translocation technology) has emerged as a predestined technique for single-molecule isolation, investigation and eventually manipulation\(^9\).

Resistive pulse sensing is a label-free, single-molecule screening technology allowing electrical detection and characterization of single particles in aqueous solutions. Thereby, location, speed and direction of motion of single target molecules can be controlled at the same time. Alternative sensing technologies include surface plasmonic/waveguide sensors, resonator-based sensors, and nanowire devices. While these technologies also offer label-free, single-molecule sensing capabilities, location control at the time of detection is low, and more sophisticated nanofabrication and integration processes are required to incorporate these types of sensors in lab-on-a-chip platforms. Instead of employing custom-labelled biological assays, resistive pulse sensing is performed purely electrically in a nanofluidic chip which contains at least one integrated nanochannel connected to two reservoirs. The system is wetted with an aqueous electrolyte solution, and a sample containing the target molecules is pumped into one reservoir. After creating a potential drop between the two reservoirs an ionic current begins to flow through the nanochannel forming the baseline current signal and creating a hydrodynamic drag force close to the channel openings. As single target molecules are pulled into the nanochannel and translocate through it one at a time, they partially block the channel and thus modulate the baseline...
current signal. This produces an individual electrical fingerprint for each translocation event whose distinct features can be correlated with the biophysical properties of the translocating molecule that created it. Resistive pulse sensing can be used to investigate a broad variety of molecular biological systems with single-particle, single nucleotide and even single-binding site resolution. It is explicitly powerful to perform next-generation nanopore-based DNA-sequencing\textsuperscript{10-16} and to investigate target-ligand interactions in the pre-clinical compound-to-hit, hit-to-lead, and lead optimization stages of the pharma pipeline\textsuperscript{17-22}.

Figure 1. Precision medicine contextualizes data from genome to exposome (left). Thereby genomic, transcriptomic, proteomic, metabolomic and microbiomic data is biomolecular by nature while epigenomic and exposomic data also takes environmental, biometric and medical metadata sources into account (center). On one hand, environmental and biometric data is highly diversified but often readily available and can be collected and processed through smart sensor networks which are incorporated into novel mobile platforms such as wearables, smart phones or smart watches. Such platforms then communicate directly to analytical tools for point-of-care monitoring and diagnostics (top). On the other hand collecting biomolecular data from intra- and intercellular systems on the genome to microbiome scale is a biophysical challenge in the field of experimental systems biology requiring the design of highly customized tools for screening biomolecules with single particle and single binding site resolution (bottom).

3. OUTLOOK

Highlighting the extraordinary current relevance of resistive pulse sensing technology for next-generation DNA-sequencing and drug discovery and giving a perspective onto the resulting impact of this technology on precision medicine applications, a comprehensive technical and commercial review has recently been published elsewhere\textsuperscript{23}. Integrated bionanosensors utilizing resistive pulse sensing technology are uniquely suitable to characterize the smallest units of the human organism and to help understand their complex interplay. While resistive pulse sensing-based next-generation sequencing platforms will mature further, the technology will also increasingly be used for developing novel diagnostics and ultra-high throughput pre-clinical drug screening systems. This could contribute to transforming the pharma pipeline similarly to how the $1000$-Genome has revolutionized DNA-sequencing, thus making resistive pulse sensing an essential part of precision medicine, the next big thing emerging at the intersection of biotechnology, nanotechnology and big data.
ACKNOWLEDGEMENTS

I thank Ajay Royyuru, Gustavo Stolovitzky, Yann Astier, Lili Deligianni, Ruhong Zhou and Binquan Luan (IBM T. J. Watson Research Center), Emmanuel Delamarche and Govind Kaigala (IBM Zurich Research Laboratory), Claudia Feger (IBM Research Brazil), Juerg v. Kaenel, John Wagner, Matthew Downton, Christine Schieber, Natalie Gunn, Sung Cheol Kim, Sridhar Kannam, Stephen Moore, Julia Baldauf (IBM Research Australia), Robert Miller (IBM Almaden Research Center), Ross Bathgate and Daniel Scott (Florey Neuroscience Institutes) as well as Stan Skafidas (University of Melbourne) for fruitful technical and strategic discussions on nanobiotechnology topics. Tom Conway and Kelly Wyres (IBM Research Australia) provided useful insights into the areas of genomics and systems biology. I also gratefully acknowledge Jonathan Rothberg (Butterfly Network), Hermann Hauser (Amadeus Capital), Stefan Roever and Roger Chen (both Genia Technologies) as well as Philip Waggoner (Ion Torrent/Thermo Fisher Scientific) for taking the time to discuss next-generation DNA-sequencing topics with me.

REFERENCES