Genet. Res., Camb. (2013), **95**, *pp.* 1–3. © Cambridge University Press 2013
doi:10.1017/S0016672313000049

EDITORIAL
Genetics research: jumping into the deep end of the pool

NOAM SHOMRON
Faculty of Medicine, Tel Aviv University, Tel Aviv 69978, Israel

Summary

The publication of the human genome, more than a decade ago, alongside the development of high-throughput technologies for DNA sequencing, marked the dawn of a new era in genetics. Large genomic projects have been initiated to decipher the mysteries hidden within the human genetic code. With the rapidly ever-growing amount of genetic information, and the importance of understanding what it all means, there is a need to generate an interdisciplinary hub that will connect researchers, both experimentalists and bioinformaticians, along with physicians and community representatives in order to develop a common genomic language. This should lead to an accessible, readable and interpretive human genome with a short list of personal actionable items. We will then be able to declare that we are moving ever closer to the point at which one's own genome will affect one's personal life at a scope beyond our current comprehension.

The publication of the human genome, more than a decade ago, marked the dawn of a new era in genetics (Lander *et al.*, 2001; Venter *et al.*, 2001). It enabled scientists to examine the genetic information from beginning to end, as a whole. However, given that it had taken many years and huge sums of money to complete the reading of one genome, it was far from feasible to further use this technology to evaluate more than a handful of additional individuals.

Professor Frederick Sanger, the British biochemist received the Nobel prize for chemistry in 1980, together with Walter Gilbert, for what was defined as 'their contributions concerning the determination of base sequences in nucleic acids'. At first, they showed how they could sequence up to 80 nucleotides per run. This was a tedious process. Nevertheless, it enabled the sequencing of more than 5000 nucleotides of the single-stranded bacteriophage φX174, the first fully sequenced genome (Sanger *et al.*, 1977). To their amazement, they were able to reveal a novel genomic feature from this one complete DNA read of multiple overlapping genes in one locus, a feature that is still being explored today (Sorek & Cossart, 2010). Subsequently, Sanger *et al.* introduced the chain-termination technique for sequencing DNA molecules, which was later known as the 'Sanger sequencing method' (Sanger *et al.*, 1980). This major leap forward in the type of sequencing approach used allowed for long stretches of DNA to be systematically and accurately recorded, laying the foundations of DNA sequencing thereof. In 1984, scientists of the Medical Research Council (MRC) in the UK were able to decode the entire DNA sequence of the Epstein–Barr virus (170 kb). Two years later, the laboratory of Leroy Hood at the California Institute of Technology (CA, USA), announced the first semi-automated DNA sequencing machine. In 1987, Applied Biosystems marketed the first automated sequencing machine that boosted sequencing such as those of human expressed sequence tags (ESTs; by Craig Ventor). Ironically, the title of one of Sanger's first sequencing papers was 'Cloning in single-stranded bacteriophage as an aid to rapid DNA sequencing' (Sanger *et al.*, 1980). Little did Sanger know that during his lifetime 'rapid' – the adjective he chose to describe 'DNA sequencing'– would take on a supersonic form.

A solution for sequencing multiple human genomes turned up about half a decade after the first human draft sequence was published. It came in the form of a second generation sequencing machine, which was also identified under the following terms: 'Next Generation Sequencing' (NGS), 'Massively Parallel

Tel: +972-3-640-6594. Fax: +972-3-640-7432. e-mail: nshomron@post.tau.ac.il

https://doi.org/10.1017/S0016672313000049 Published online by Cambridge University Press

Sequencing' (MPS), 'High Throughput Sequencing' (HTS), or 'Deep Sequencing'. This revolutionary technology enabled reading an individual human genome in a matter of days (Shendure & Lieberman Aiden, 2012). This notion made grand projects realistic, such as the sequencing of: (i) 1000 genomes from 13 different populations (1000 Genomes Project Consortium, 2012); (ii) thousands of cancer genomes (Boehm & Hahn, 2011); and (iii) the entire microorganisms of a human gut (Human Microbiome Project Consortium, 2012).

The technology behind the first and second generation sequencers is conceptually similar. Fragments of unknown DNA are flanked by linkers, amplified (in most cases) and read by consecutive light emission (or change of chemical state) once a complementary nucleotide is incorporated (A-T and G-C). However, in the second generation apparatuses, this process occurs around 10–100 million times per experiment representing a 1–10 million fold increase in read depth in just over three decades. For comparison, and in order to grasp this meteoric advance, think about the magnification of a light microscope compared with an electron microscope: about $\times 400$ versus $\times 200\,000$; a 50-fold increase. Similarly, in the transportation world, the progress made from one of the first cars (more than 100 years ago) which cruised at 4 km per hour, versus a space shuttle accelerating to leave planet Earth at about 40 000 km per hour, presents a mere 10 000-fold increase. Thus, these exhilarating abilities will no doubt advance current research and will evidently progress our profound understanding of the human genome.

Currently, there are numerous scientific laboratories and companies that utilize massive sequencing for the study of human genetics on a regular basis. These projects either record the base composition of the entire 23 human chromosomes or focus on reading all protein-coding regions in the human genome, also known as the 'Exome'. In the near future, it is safe to assume that every individual would carry their own genetic makeup on a digital media device.

In order to interpret the information stored in the DNA, powerful bioinformatics analysis must be implemented by the sequencing team. By means of computational investigation, scientists attempt to link particular genetic composition, or changes thereof, to functional outcomes or phenotypes (Isakov & Shomron, 2011). The advent of genome sequencing allows for the: (i) identification of genetic diseases (Walsh *et al.*, 2010; Fuchs-Telem *et al.*, 2012); (ii) mapping of cancerous tissue (Ley *et al.*, 2008); and (iii) profiling of pathogen infections (Isakov *et al.*, 2011), to name a few. A decade ago, fewer than 100 genetic disease-causative genes were identified. Today, nearly 3000 Mendelian diseased genes have been revealed, and the list is rapidly growing with the increase in the number of genetic and physical maps created by every genome sequenced.

For scientific researchers, receiving the complete DNA sequence of an organism is as straightforward as supplying them with a substrate to work on. It is similar to allowing a mechanic to look at the car's blueprints before attempting to fix it. For physicians, the comprehensive view of the DNA allows an unbiased examination of genomic information, which serves as a possible link to the clinical evaluation and treatment management. Some physicians describe the interpretation of a patient's genetic makeup as a 'gift' that enables them to look 'outside the box' and explore the genetic causes of symptoms, which they would have never done otherwise. For the general public, access to one's own genetic profile currently opens a Pandora box with a myriad of questions and very few answers. This will soon change owing to the intensive research this new technology enables.

With the rapidly ever-growing amount of genetic information, and the importance of understanding what it all means, there is a need to generate an interdisciplinary hub that will connect researchers, both experimentalists and bioinformaticians, along with physicians and community representatives in order to come up with a common genomic language. This should lead to an accessible, readable and interpretive human genome with a short list of personal actionable items. We will then be able to declare that we are moving ever closer to the point at which one's own genome will affect one's personal life at a scope beyond our current comprehension.

All of the above sums up in essence what *Genetics Research* hopes to achieve going forward and it will become the forum where these new and exciting challenges will be highlighted, debated and disseminated. I look forward to welcoming this groundbreaking research to the journal!

Acknowledgements

I thank the Shomron laboratory for their valuable discussions and comments on the manuscript. The Shomron laboratory is supported by the Wolfson family Charitable Fund; Claire and Amedee Maratier Institute for the Study of Blindness and Visual Disorders; Israel, Frida and Haya Hamer Fellowship; Levine Katan Leukemia Research Fellowship; Kurz-Lion Foundation; I-CORE Program of the Planning and Budgeting Committee and the Israel Science Foundation [Grant No. 41/11].

References

1000 Genomes Project Consortium, Abecasis, G. R., Auton, A., Brooks, L. D., DePristo, M. A., Durbin, R. M., Handsaker, R. E., Kang, H. M., Marth, G. T. & McVean, G. A. (2012). An integrated map of genetic variation from 1,092 human genomes. *Nature* **491**, 56–65.

https://doi.org/10.1017/S0016672313000049 Published online by Cambridge University Press

Boehm, J. S. & Hahn, W. C. (2011). Towards systematic functional characterization of cancer genomes. *Nature Reviews Genetics* **12**, 487–498.

Fuchs-Telem, D., Sarig, O., van Steensel, M. A., Isakov, O., Israeli, S., Nousbeck, J., Richard, K., Winnepenninckx, V., Vernooij, M., Shomron, N., Uitto, J., Fleckman, P., Richard, G. & Sprecher, E. (2012). Familial pityriasis rubra pilaris is caused by mutations in CARD14. *American Journal of Human Genetics* **91**, 163–170.

Human Microbiome Project Consortium (2012). A framework for human microbiome research. *Nature* **486**, 215–221.

Isakov, O. & Shomron, N. (2011). Deep sequencing data analysis: challenges and solutions. In *Bioinformatics – Trends and Methodologies* (ed. M. A. Mahdavi). Rijeka, Croatia: InTech. Available from: http://www.intechopen.com/books/bioinformatics-trends-and-methodologies/deepsequencing-data-analysis-challenges-and-solutions.

Isakov, O., Modai, S. & Shomron, N. (2011). Pathogen detection using short-RNA deep sequencing subtraction and assembly. *Bioinformatics* **27**, 2027–2030.

Lander, E. S., Linton, L. M., Birren, B., Nusbaum, C., Zody, M. C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., Funke, R., Gage, D. *et al*. (2001). Initial sequencing and analysis of the human genome. *Nature* **409**, 860–921.

Ley, T. J., Mardis, E. R., Ding, L., Fulton, B., McLellan, M. D., Chen, K., Dooling, D., Dunford-Shore, B. H., McGrath, S., Hickenbotham, M., Cook, L., Abbott, R. *et al*. (2008). DNA sequencing of a cytogenetically normal acute myeloid leukaemia genome. *Nature* **456**, 66–72.

Sanger, F., Air, G. M., Barrell, B. G., Brown, N. L., Coulson, A. R., Fiddes, C. A., Hutchinson, C. A., Slocombe, P. M. & Smith, M. (1977). Nucleotide sequence of bacteriophage φX174 DNA. *Nature* **265**, 687–695.

Sanger, F., Coulson, A. R., Barrell, B. G., Smith, A. J. & Roe, B. A. (1980). Cloning in single-stranded bacteriophage as an aid to rapid DNA sequencing. *Journal of Molecular Biology* **143**, 161–178.

Shendure, J. & Lieberman Aiden, E. (2012). The expanding scope of DNA sequencing. *Nature Biotechnology* **30**, 1084–1094.

Sorek, R. & Cossart, P. (2010). Prokaryotic transcriptomics: a new view on regulation, physiology and pathogenicity. *Nature Reviews Genetics* **11**, 9–16.

Venter, J. C., Adams, M. D., Myers, E. W., Li, P. W., Mural, R. J., Sutton, G. G., Smith, H. O., Yandell, M., Evans, C. A., Holt, R. A., Gocayne, J. D., Amanatides, P. *et al*. (2001). The sequence of the human genome. *Science* **291**, 1304–1351.

Walsh, T., Shahin, H., Elkan-Miller, T., Lee, M. K., Thornton, A. M., Roeb, W., Abu Rayyan, A., Loulus, S., Avraham, K. B., King, M. C. & Kanaan, M. (2010). Whole exome sequencing and homozygosity mapping identify mutation in the cell polarity protein GPSM2 as the cause of nonsyndromic hearing loss DFNB82. *American Journal of Human Genetics* **87**, 90–94.

https://doi.org/10.1017/S0016672313000049 Published online by Cambridge University Press