

Sequencing your genome: your future is here, but are you sure you want to know it?

NIR PILLAR¹, OFER ISAKOV¹ AND NOAM SHOMRON^{1*}

¹Faculty of Medicine, Tel Aviv University, Tel Aviv 69978, Israel

(Received 18 March 2014; accepted 21 May 2014)

Next-generation sequencing (NGS; also known as deep sequencing or ultra-high throughput sequencing) has probably been the most important tool for genomic research over the past few years. NGS has led to numerous discoveries and scientific breakthroughs in the genetic field. The sequencing technology that has entered the research laboratory in the past decade is now being introduced into the clinical diagnostic laboratory. Consequently, NGS results are becoming available in the medical arena as abundance of clinically relevant variants, conferring predisposition to disease, are being discovered at a growing rate (Stanley, 2014).

Since its first utilization, NGS has been revolutionizing nearly all medical disciplines. Examples include *de novo* mutation studies (Soler, 2013), cancer genomics discovery (Shern, 2014), prediction of targeted therapy/actionable mutations (Kohsaka, 2014) and noninvasive prenatal testing (Tsui, 2014). The implementation and integration of NGS in these fields is ascertained by the flood of relevant publications (close to 1,500 NCBI PubMed records, under the term 'next generation sequencing' or 'deep sequencing' in humans, during 2013). This immense inflation of NGS-derived data, its accessibility, its transfer to the medical field and the overall interest means that there is a need for an overview on its implications on the medical arena and the general public.

NGS allows sequencing of an entire human genome (whole-genome sequencing; WGS) or the entire protein-coding DNA, representing 1–3% of the human genome (whole-exome sequencing; WES) in less than a day. This allows for large-scale, comprehensive studies comprising many sequenced individuals to be performed with unprecedented speed.

Decreased sequencing costs (with one major sequencing provider, Illumina, claiming the arrival of

the \$1000 genome (Hayden, 2014)) have provided remarkable opportunities for the launch of large collaborative projects focused on WGS. As more people have their genome sequenced, the interpretation and understanding of human genetic variance becomes deeper. This in turn facilitates the differentiation between pathogenic variants and variants of uncertain significance. Moreover, this accumulation of population genetic data potentially increases the number of identified clinically relevant variants that are medically actionable.

Certain points have to be considered while debating if genome sequencing should be available to the general public and be performed regardless of clinical indication, whether it is following a physician's recommendation or through a private setting, possibly marketed as an 'X-ray for the genome'. When facing the decision whether to sequence your own genome, there are several arguments supporting and opposing this decision.

Advantages and implementations:

- Development of an effective preventive medicine platform by collecting the entire set of an individual's genetic variations may uncover certain known, disease-associated mutations. Some of these diseases may be alleviated by lifestyle modifications and/or early medical treatment. Early detection of these types of mutations can prevent morbidity and might delay mortality. In a recent report, 42% of people surveyed introduced positive changes into their health behavior after receiving genomics results (Green, 2014). Four pilot research programs were recently launched by the National Institutes of Health (NIH) to explore the potential of newborn genome sequencing in preventing a vast array of disorders and conditions starting at birth (NIH, 2013).
- Precision medicine, also known as pharmacogenomics, allows for the elucidation of the effect that specific genetic variants have on the metabolism and action of a certain drug. This field is

* Corresponding author: Tel: +972-3-6406594. Fax: +972-3-6407432. E-mail: nshomron@post.tau.ac.il

developing rapidly, with a constant increase in known mutations modulating drug response. The field of personalized medicine was initiated by revealing single nucleotide polymorphisms (SNPs) that account for some of the genetic variability between individuals and made possible the use of genome-wide association studies (GWAS) to examine genetic variation and risk for many common diseases. One example is the clinical impact of the major CYP polymorphisms (CYP2D6, CYP2C19 and CYP2C9) on drug therapy. CYP2D6 is responsible for the oxidative metabolism of about 25% of prescribed drugs (NIH, 2013). The ultra-rapid metabolizer phenotype can cause therapeutic inefficacy of antidepressants, whereas an increased risk of toxicity has been reported in poor metabolizers with several psychotropics (Samer, 2013). Variations in CYP2C19 are associated with decreased clopidogrel metabolic activation, resulting in a decreased antiplatelet effect, and an increased likelihood of a cardiovascular event (Mega, 2009). Besides SNP detection, NGS enables detection of many other genetic variations—such as small insertions and deletions ranging from 1 to 10,000 bp in length and larger forms of structural variation that may optimize patient drug treatment by tailoring specific pharmacogenomics-based therapy and better monitoring of genomic biomarkers for each individual patient.

- While the combined costs of sequencing, data interpretation and storage are still high, they represent a one-time investment for the health provider. Personal sequences remain intact from cradle to grave and a single sequencing effort may aid in countless health events during a person's lifetime.
- The public has a fundamental right to know, collect or receive information about their own health, which also includes their genome content, whether it involves positive or negative outcomes. If any government, social or religious agency/group prohibits this privilege, it would require strong legal or moral justification.
- 'Democratizing' health care, expanding the access given to individuals and enabling the acquirement of as much data as they can in order to make the most informative and founded choices concerning their own health. This follows the historical trend of patient empowerment, a shift from a paternalistic approach to patient care, which tends to ignore personal preferences, towards a patient-centered care in which one's ability to understand and manage his or her own health and disease is improved (No authors listed, 2012). Empowered patients have a better understanding of how to navigate between the many players in the health care system including family, physicians, health insurers and health care regulators. Coping with the complexities of health

systems is crucial for achieving better health outcomes. Access to as much information as possible better equips the public to handle this task.

- By having an individual's genome sequenced, upon consent, information may be placed in a database, which can be shared by researchers and geneticists around the world. This data can then be used for the development of tools and interpretations based on the genomic information. Unsustainable cost increases threaten the global health care system, and additional progress is stymied more by social than technological factors. Many algorithms have been developed to detect and characterize complex genetic rearrangements, such as insertions, deletions, duplications and inversions, in the human genome. However, these computational tools suffer from relatively high false-positive and false-negative rates compared to SNPs. One suggestion is that larger cohorts are pivotal to improving these success rates. Only by engaging health care consumers (patients) as pioneers who provide both health-related data and insights into pathophysiology can we meet these societal challenges and thus accelerate the pace of biomedical innovation (Hood, 2014).

Current challenges and limitations:

- Despite the fact that sequencing costs are rapidly declining, the price for computational analysis and data storage remain high, and may pose a financial burden on health care providers and institutions if it becomes a routine procedure.
- Although genetic data are considered protected health information under the US Health Insurance Portability and Accountability Act (HIPAA) and similar laws worldwide, many of the protections disappear when the information is 'de-identified', meaning that the 18 identifiers specified in the act (including names, addresses, birthdates etc.) are removed (Angrist, 2013). Cross-reference information from public databases to discover the identity of DNA samples donated to research has recently been proven (Gymrek, 2013) and might cause any genome-sequenced individual to unintentionally reveal personal information. The risk of re-identification from genomic data sources may affect not only the sequenced individuals, but also their first-degree relatives with whom they share a large portion of their genome sequence. Moreover, we now live in a 'global village' – people often move between countries and continents. However there is no global privacy act among the various countries or health providers. While a person might reside in a country with outstanding genetic data protection, they may later move on to countries with no such protections, where leakage of their personal genome data may take place.

- A huge gap still exists between the total amount of genomic data resulting from NGS and the interpretable amount according to the current knowledge base. The lack of informed clinicians and geneticists increases this gap, making it challenging for the patient to interpret their own genetic information. The data interpretation gap is demonstrated by the case of 23andMe, a firm offering SNP genotyping services directly to consumers, which was limited by the US Food and Drug Administration (FDA) in December 2013 in order to avoid the pitfalls that can arise from false-positives or false-negatives for certain gene variants. Here we have a ‘catch 22’ situation: this gap cannot be closed unless many thousands of people will contribute their personal genomes along with their health records. The Personal Genome Project (PGP) approach has begun to address this issue by limiting any promises of privacy and confidentiality (Church, 2005). The project, founded in 2005, is dedicated to creating a shared public platform of public genome, health and trait data. To date, it has more than 2,000 participants, all of whom have agreed to make public, and potentially identifiable, any genomic, medical, environmental and trait data collected about them during the study in order to help science understand the relationship between genotype and phenotype.
- Sequencing individuals’ DNA and informing them that they have a slightly elevated genetic risk for a certain disease could result in great disservice. This information might motivate them to behave insensibly. This in turn may uncover marketing opportunities for the private sector that will promote all sorts of clinically unnecessary and potentially damaging screening, with further negative and unintended consequences. More than a decade since the Human Genome Project was declared complete, fewer than 60 genetic variants are deemed worthy for use in clinical care, most for severe conditions in very young children (Green, 2013). Weeding out irrelevant mutations, focusing on the significant ones and deciding how to act accordingly is a tough mission for health professionals, let alone for the general public.
- There would be a surge in demand for all kinds of tests, most of these would be unnecessary and would create a significant burden on the health care system. A recent study claimed that more than a quarter of complex molecular genetic testing orders are unnecessary (Miller, 2014) and this number is expected to rise as more patients demand further validations for their genome-sequencing analysis.
- Some patients prefer not to know their genetic susceptibilities, especially if there is no cure for these mutated genes. One survey showed that 50–70%

Table 1. Considerations one should take prior to personal genome sequencing

Advantages and implementations	Challenges and limitations
Preventive medicine improvement	Still expensive
Personalized treatment development	Privacy breach possibilities
A one-time investment for life	Large interpretation gap
The right to know your genetic data	Motivation to behave insensibly
‘Democratizing’ health care	Surge of unnecessary tests
Shared information will advance health	Lack of Information

of patients would like to be informed about all their results from genomic studies while 16–20% considered genetic information to be riskier than other types of personal medical data (Ruiz-Canela, 2011). Another study emphasized that the majority of patients desire to have the choice to learn about the probability of conditions that may develop, but given the choice, are not sure how they would act (Townsend, 2012).

- The finding that many putatively damaging variants are present as homozygotes in the genomes of healthy individuals raises questions concerning the extent to which these variants contribute to the determination of phenotype. Several mutations reported as ‘disease-causing mutations’ in the Human Gene Mutation Database (HGMD) are likely to be annotation errors (Xue, 2012). Thus, it has to be carefully considered that NGS projects involving healthy people may result in the incidental discovery of such silent ‘disease-related’ variants (Table 1).

Recently, the American College of Medical Genetics (ACMG) has published a recommendation that laboratories conducting clinical sequencing of patients should report mutations in 56 genes associated with highly penetrant, monogenic disorders regardless of the indication for which the sequencing was ordered, and made the patient’s clinician or care team responsible for consulting the patient regarding these results (Green, 2013). It is reasonable to speculate that in the foreseeable future, as the cost of such sequencing continues to fall, a majority of health plans will make it easy for their members to have their entire genomes sequenced and linked to their electronic health records for these 56 genes and many others to come. Millions of people will probably be keen to have their genomes sequenced and may even be willing to pay for the sequencing themselves. The challenging part will be, as is today, the clinical interpretation of those genomes (Isakov, 2013) and the development of useful

patient–consumer recommendations. An analogy to radiology, where unexpected anomalies are frequently found, is commonly used to explain this challenging complexity (Solomon, 2014). However, despite impressive ongoing efforts that will continue to yield great progress in the near future, we are not at the point where interpreting a genomic data set is similar to interpreting a radiologic study. It is important not to oversell our abilities to understand and manage the complexity of the human genome in clinical practice.

Though historically human genetic studies focused on the rare and unusual, and was dominated by a handful of experts, the new wave of understanding of the human genome cannot happen without our large-scale participation. Dealing with the inevitable limits of knowledge in a constructive manner is fundamentally different from setting limits to data generation in the presumed ‘best interest’ of the patient or as a method to mask uncertainty and limit the doctor’s liability (Lunshof, 2014; Jarvik, 2014). However, when there is still so much about the function of genes to be uncovered, how useful will free access to DNA sequencing be? And when we reach a point that enough information is gained and interpreted, would you be ready to read your genes in order to know your own medical future?

We thank Dr David Gurwitz, Prof Karen Avraham and Dr Elena Milanesi for commenting on our manuscript.

Funding

The Shomron laboratory is supported by the Israel Cancer Research Fund (ICRF), Research Career Development Award (RCDA); Wolfson Family Charitable Fund; Earlier.org – Friends for an Earlier Breast Cancer Test; Claire and Amedee Maratier Institute for the Study of Blindness and Visual Disorders; I-CORE Program of the Planning and Budgeting Committee, The Israel Science Foundation (grant number 41/11); The Israeli Ministry of Defense, Office of Assistant Minister of Defense for Chemical, Biological, Radiological and Nuclear (CBRN) Defense; Foundation Fighting Blindness; and the Saban Family Foundation – Melanoma Research Alliance.

References

- Angrist, M. (2013). Genetic privacy needs a more nuanced approach. *Nature* **494**, 7.
- Church, G.M. (2005). The personal genome project. *Molecular Systems Biology* **1**, 2005.0030.
- Green, R.C. & Farahany, N.A. (2014). Regulation: the FDA is overcautious on consumer genomics. *Nature* **505**, 286–287.
- Green, R.C., Berg, J.S., Grody, W.W., Kalia, S.S., Korf, B.R., Martin, C.L., McGuire, A.L., Nussbaum, R.L., O’Daniel, J.M., Ormond, K.E., Rehm, H.L., Watson, M.S., Williams, M.S., Biasecker, L.G. & American College of Medical Genetics and Genomics (2013). ACMG recommendations for reporting of incidental findings in clinical exome and genome sequencing. *Genetics in Medicine* **15**, 565–574.
- Gymrek, M., McGuire, A.L., Golan, D., Halperin, E. & Erlich, Y. (2013). Identifying personal genomes by surname inference. *Science* **339**, 321–324.
- Hayden, E.C. (2014). Is the \$1,000 genome for real? *Nature News*. (<http://www.nature.com/news/is-the-1-000-genome-for-real-1.14530>). Accessed 15 January 2014.
- Hood, L. & Price, N.D. (2014). Demystifying disease, democratizing health care. *Science Translational Medicine* **6**, 225ed5.
- Isakov, O., Perrone, M. & Shomron, N. (2013). Exome sequencing analysis: a guide to disease variant detection. *Methods in Molecular Biology* **1038**, 137–158.
- Jarvik, G.P., Amendola, L.M., Berg, J.S., Brothers, K., Clayton, E.W., Chung, W., Evans, B.J., Evans, J.P., Fullerton, S.M., Gallego, C.J., Garrison, N.A., Gray, S.W., Holm, I.A., Kullo, I.J., Lehmann, L.S., McCarty, C., Prows, C.A., Rehm, H.L., Sharp, R.R., Salama, J., Sanderson, S., Van Driest, S.L., Williams, M.S., Wolf, S.M., Wolf, W.A., eMERGE Act-ROR Committee and CERC Committee, CSER Act-ROR Working Group & Burke, W. (2014). Return of genomic results to research participants: the floor, the ceiling, and the choices in between. *American Journal of Human Genetics* **pii**, S0002-9297(14)00181-5.
- Kohsaka, S., Shukla, N., Ameer, N., Ito, T., Ng, C.K., Wang, L., Lim, D., Marchetti, A., Viale, A., Pirun, M., Socci, N.D., Qin, L.X., Sciot, R., Bridge, J., Singer, S., Meyers, P., Wexler, L.H., Barr, F.G., Dogan, S., Fletcher, J.A., Reis-Filho, J.S. & Ladanyi, M. (2014). A recurrent neomorphic mutation in *MYO1D* defines a clinically aggressive subset of embryonal rhabdomyosarcoma associated with PI3K-AKT pathway mutations. *Nature Genetics* **46**, 595–600.
- Lunshof, J.E., Church, G.M. & Prainsack, B. (2014). Information access. Raw personal data: providing access. *Science* **343**, 373–374.
- Mega, J.L., Close, S.L., Wiviott, S.D., Shen, L., Hockett, R.D., Brandt, J.T., Walker, J.R., Antman, E.M., Macias, W., Braunwald, E. & Sabatine, M.S. (2009). Cytochrome p-450 polymorphisms and response to clopidogrel. *The New England Journal of Medicine* **360**, 354–362.
- Miller, C.E., Krautscheid, P., Baldwin, E.E., Tvrđik, T., Openshaw, A.S., Hart, K. & Lagrave, D. (2014). Genetic counselor review of genetic test orders in a reference laboratory reduces unnecessary testing. *American Journal of Medical Genetics, Part A* **164**, 1094–1101.
- NIH program explores the use of genomic sequencing in newborn healthcare. NIH Website (2013). (<http://www.nih.gov/news/health/sep2013/nhgri-04.htm>). Accessed 4 September 2013.
- No authors listed. (2012). Patient empowerment—who empowers whom? *Lancet* **379**, 1677.
- Ruiz-Canela, M., Valle-Mansilla, J.I. & Sulmasy, D.P. (2011). What research participants want to know about genetic research results: the impact of “genetic exceptionalism”. *Journal of Empirical Research on Human Research Ethics* **6**, 39–46.
- Samer, C.F., Lorenzini, K.I., Rollason, V., Daali, Y. & Desmeules, J.A. (2013). Applications of CYP450 testing

- in the clinical setting. *Molecular Diagnosis and Therapy* **17**, 165–184.
- Shern, J. F., Chen, L., Chmielecki, J., Wei, J. S., Patidar, R., Rosenberg, M., Ambrogio, L., Auclair, D., Wang, J., Song, Y. K., Tolman, C., Hurd, L., Liao, H., Zhang, S., Bogen, D., Brohl, A. S., Sindiri, S., Catchpoole, D., Badgett, T., Getz, G., Mora, J., Anderson, J. R., Skapek, S. X., Barr, F. G., Meyerson, M., Hawkins, D. S. & Khan, J. (2014). Comprehensive genomic analysis of rhabdomyosarcoma reveals a landscape of alterations affecting a common genetic axis in fusion-positive and fusion-negative tumors. *Cancer Discovery* **4**, 216–231.
- Soler, V. J., Tran-Viet, K. N., Galiacy, S. D., Limviphuvadh, V., Klemm, T. P., St Germain, E., Fournié, P. R., Guillaud, C., Maurer-Stroh, S., Hawthorne, F., Suarez, C., Kantelip, B., Afshari, N. A., Creveaux, I., Luo, X., Meng, W., Calvas, P., Cassagne, M., Arné, J. L., Rozen, S. G., Malecaze, F. & Young, T. L. (2013). Whole exome sequencing identifies a mutation for a novel form of corneal intraepithelial dyskeratosis. *Journal of Medical Genetics* **50**, 246–254.
- Solomon, B. D. (2014). Incidentalomas in genomics and radiology. *The New England Journal of Medicine* **370**, 988–990.
- Stanley, C. M., Sunyaev, S. R., Greenblatt, M. S. & Oetting, W. S. (2014). Clinically Relevant variants – identifying, collecting, interpreting, and disseminating: the 2013 Annual Scientific Meeting of the Human Genome Variation Society. *Human Mutation* **35**, 505–510.
- Townsend, A., Adam, S., Birch, P. H., Lohn, Z., Rousseau, F. & Friedman, J. M. (2012). “I want to know what’s in Pandora’s Box”: comparing stakeholder perspectives on incidental findings in clinical whole genomic sequencing. *American Journal of Medical Genetics, Part A* **158A**, 2519–2525.
- Tsui, N. B., Jiang, P., Wong, Y. F., Leung, T. Y., Chan, K. C., Chiu, R. W., Sun, H. & Lo, Y. M. (2014). Maternal plasma RNA sequencing for genomewide transcriptomic profiling and identification of pregnancy-associated transcripts. *Clinical Chemistry* [Epub ahead of print].
- Xue, Y., Chen, Y., Ayub, Q., Huang, N., Ball, E. V., Mort, M., Phillips, A. D., Shaw, K., Stenson, P. D., Cooper, D. N., Tyler-Smith, C. & 1000 Genomes Project Consortium (2012). Deleterious- and disease-allele prevalence in healthy individuals: insights from current predictions, mutation databases, and population-scale resequencing. *American Journal of Human Genetics* **91**, 1022–1032.