

1

Next-Generation Sequencing for Clinical Diagnostics of Cardiomyopathies

Jan Haas, Hugo A. Katus, and Benjamin Meder

1.1

Introduction

The vast progress next-generation sequencing (NGS) has undergone during the past few years [1,2] has opened doors for a more advanced genetic diagnostic for many inherited diseases, such as Miller syndrome or Charcot–Marie–Tooth neuropathy [3,4]. Here, we want to describe the paradigm change in genetic diagnostics using the example of cardiomyopathies.

1.2

Cardiomyopathies and Why Genetic Testing is Needed

Cardiomyopathies are a heterogeneous group of cardiac diseases that can either be acquired through, for example, inflammation (myocarditis), be stress-induced (tako-tsubo), or be due to a genetic cause [5,6]. Examples of genetic forms are hypertrophic cardiomyopathy (HCM), dilated cardiomyopathy (DCM), arrhythmogenic right ventricular cardiomyopathy, and left-ventricular non-compaction cardiomyopathy. Together with the channelopathies, such as long-QT syndrome and Brugada syndrome, they account for the most common heart diseases and belong to the most prevalent causes of premature death in western civilizations [7,8]. A point mutation in exon 13 of the β -myosin heavy chain gene was the first detected mutation diagnosed to be relevant for HCM in 1990 [9]. Driven by this finding, genetic research has progressed tremendously over the past two decades. Mutations in genes coding for a diverse set of proteins (e.g., sarcomeric, cytoskeletal, desmosomal, channel and channel-associated, membrane, and nuclear proteins, but also mitochondrial proteins or proteins relevant for mRNA splicing) have now been found to be implicated in disease onset and progression [10]. With currently more than 90 known disease genes with more than 1000 exons and multiple malign mutations per gene, the disease's heterogeneity is high and poses a challenge for classical Sanger-based sequencing. Although Sanger sequencing is able to detect mutations by testing only the

most heavily affected genes, such as the β -myosin heavy chain gene (*MYH7*), where it is possible to find mutations in up to 30–50% of HCM patients, the mutation frequency in most genes is very low [10–12]. Therefore, new methods were needed to further improve genetic diagnostics in cardiomyopathy patients.

1.3 NGS

In contrast to Sanger sequencing, which is only capable of sequencing a few megabases, NGS is able to sequence hundreds of gigabases per run [13–15]. Currently, the most widely used NGS systems are the “sequencing by synthesis”-based sequencer HiSeq 2000 (Illumina), the “ligation and two-base coding”-based system SOLiDv4 (Life technologies), and the 454 GS FLX (Roche), which relies on “pyrosequencing” technology [7,16]. A detailed comparison of currently used systems including performance benchmarks, such as read lengths and output amounts, has been published recently by Liu *et al.* [1,2,17]. In addition to the mature NGS systems, so-called benchtop sequencers have emerged. Those instruments, such as the MiSeq (Illumina), the Ion Torrent (PGM), or the GS Junior (Roche), benefit from a significantly shorter run time (hours compared to days) and a lower price, taking into account a reduced amount of sequenced bases [17,18].

Originally, NGS was designed to sequence whole genomes. In order to reduce costs, methods for target enrichment were developed to restrict sequencing to the regions of interest only [19–22]. The sequencing of only selected segments of the genome allowed the sequencing of a larger number of individuals per run [23]. For the enrichment of the target regions, array-based, in-solution based, or polymerase chain reaction (PCR)-based approaches exist [7,24]. Depending on the sequence composition of the target region in terms of, for example, GC content or sequence heterogeneity, the efficiency of the different methods might vary [25]. In recent years, the number of NGS applications has grown considerably. In addition to the target-enrichment methods, which can be either used for custom gene panels or whole-exome sequencing (WES), RNA-Seq methods to study messenger RNA as well as microRNA are now also routinely used. Furthermore, Methyl-Seq or Chip-Seq methods to study DNA methylation are frequently used. It can be assumed that the number of applications will grow further together with the need for more advanced bioinformatics analysis tools [26]. Here, a shift in the cost distribution from sequencing to downstream analysis is expected [27].

1.4 NGS for Cardiomyopathies

As mentioned above, whole-genome sequencing (WGS) is an unbiased approach to determine the exact order of every base in a studied genome. In the case of patient studies, the human genome, consisting of more than 3 billion bases,

needs to be analyzed. Although costs have been decreasing dramatically (<http://www.genome.gov/sequencingcosts>), it is still too expensive to perform WGS with a sufficient coverage for routine diagnostics in cardiomyopathies. Downstream bioinformatics analyses are also much more demanding for WGS compared with WES or partial-exome sequencing (PES), which is preferred to be used instead. Whereas WES is mainly used to discover new disease genes, PES has started to become a standard approach for high-throughput testing of multiple genes and patients. Meder and Haas, for example, applied an array-based enrichment of 47 genes (0.27 Mb) on cardiomyopathy patients, and were able to identify disease-causing mutations in both HCM (80%) and DCM (40%) patients [28]. A similar smaller-scaled array-based approach was used by Mook *et al.* to study 23 genes in cardiomyopathy patients [29]. PCR-based, filter-based, and in-solution-based methods have also successfully been applied in combination with the different NGS systems mentioned above to study cardiomyopathy-relevant genomic loci by NGS [30–34]. Haas *et al.* for example were able to for the first time show the mutational landscape for DCM across a large European cohort of 639 patients using in solution based target enrichment (Haas *et al.* EUR HEAR J. 2014). These studies show the feasibility of using NGS in a clinical environment, but also show the diversity of currently used approaches. Although studies exist which claim NGS to be ready as a stand-alone diagnostic test [35], most centers still rely on Sanger validation of the relevant NGS variants, which are still mainly produced within research projects and are not yet part of the daily routine. Although initial guidelines for clinical NGS testing exist [36,37], it is still difficult to compare results between different centers.

1.5

Sample Preparation

Well-established protocols exist for sample preparation that enable technicians to reproducibly prepare high-quality sequencing libraries, if high-quality DNA is used. Therefore, an initial quality check of the DNA through, for example, Bioanalyzer (Agilent) or Qubit (Invitrogen) measurements is inevitable. Library preparation protocols have been improved tremendously and now only require a few nanograms of input DNA compared with the several micrograms that were needed not so long ago. Also, the time needed for sample preparation has been shortened from a couple of days to a few hours, making it possible to finish preparation in a single working day, achieved, for example, with the Haloplex system (Agilent). Recently, disease-specific enrichment assays were introduced by Haloplex, including an optimized predesigned cardiomyopathy and arrhythmia panel, removing the need for manual target region design. Such panels already exist for cancer, for example, and are expected to be developed for other diseases. Another important aspect in the course of sample preparation is the tracking of the samples to guarantee sample integrity when used in a high-throughput manner. Use of a laboratory information management system

(LIMS) is desired. Here, freely as well as commercially available tools exist that help to reduce manual intervention and lower the overall turnaround time [38,39].

1.6

Bioinformatics Analysis Pipeline

Nowadays, the decreasing costs per base enable researchers to sequence larger target regions, exomes or genomes at a higher depth. However, those high-quality gigabase-scale datasets pose an immense challenge for downstream analyses. One major problem for comparison of results is the variety of mainly “homegrown” analysis strategies that have been developed at individual sites. Briefly, they consist of mapping, variant calling, annotation, filtering, and validation of selected variants. Although most approaches rely on similar strategies for filtering (e.g., filtering variants present in databases like dbSNP or the 1000 Genome Project) caution has to be taken. Andreasen *et al.* showed, for example, that 14% of HCM and 17% of DCM previously disease-causing reported missense and nonsense variants are present within the National Heart, Lung and Blood Institute “Grand Opportunity” Exome Sequencing Project (GO-ESP) cohort, which contains exome data from 6500 individuals [40]. Depending on the chosen filters and also due to differences in pipeline tools, tool combinations, or even versions of the programs, this will lead to a low concordance among the analyses [41]. Despite those drawbacks, it is expected that these hurdles will be overcome by newly developed, improved algorithms. Growing sequencing quality and performance of analysis tools will contribute to provide reliable variant calls, with no need for validation to gain acceptable clinical sensitivity, specificity, and positive (negative) predictive values ready for clinical use in the near future. Such a “routine use” requires an existing infrastructure in both the wet-lab and bioinformatics.

Overall, costs are still high to fully equip a laboratory for NGS-based genetic testing [1]. Benchtop sequencers may become a cheaper solution for some diseases, depending on the required sequencing capacity to adequately cover the target region. Due to the discussed differences in NGS techniques and analysis tools, no ultimate cutoff rule for base coverage exists. However a minimum coverage of 30 times seems to provide genuine results and is applied as a standard for many laboratories [36]. Researchers have to calculate the necessary sequencing capacity for their desired coverage based on their target region and decide which NGS system fits best.

1.7

Interpretation of Results and Translation into Clinical Practice

Apart from any limitation mentioned above, the translation of NGS into daily genetic testing of cardiomyopathies is mainly hindered by the restraints

physicians have in interpretation of the finally reported annotated variants. Disease mutation databases like the Human Genome Mutation Database (<http://www.biobase-international.com/product/hgmd>) help to identify known mutations and can give a hint to their contribution to disease onset or progression. However, for cardiomyopathies, it is expected that many cases (sporadic and familial) are caused by very rare or private variants. Thus, finding a new disease mutation is like finding a needle in a haystack. Whereas nonsense mutations in a known disease gene are expected to cause a disease, others may be reported as “variants of unknown significance” and need to be investigated further. One way of dissecting the possible influence a variant has on protein function is to apply prediction algorithms. A variety of stand-alone as well as Web-based tools exist [42–50]. Their calculations are based on, for example, conservation of the amino acid, transition frequencies, domain profiles, structural positions, or post-translational modifications. Most of them are trained with certain sets of variants. Depending on the type of tested amino acid (e.g., charging state), the tools all perform differently in terms of sensitivity [51,52]. The degree of uniformity among the tools also differs and some tools tend to predict much more damaging variants than others [53]. Despite of the limitations, these tools are valuable to filter out possibly benign variants and restrict validation to only a smaller candidate set, which should be studied in more depth. Such a further investigation comprises, for example, segregation analysis and functional assays in animal models. For cardiomyopathies, the zebrafish has been proven to be an excellent model organism to test a gene/mutation function by knockdown or overexpression analyses [54]. We have also used the zebrafish in the previously mentioned study to validate novel disease variants. With this approach we were able to prove the malignancy of a variant on zebrafish heart function together with a co-segregation in the index patient’s family [28]. Another important method to reveal a variant’s effect is genotype–phenotype correlational analyses. Lopes *et al.* used such an approach to compare rare non-synonymous single nucleotide polymorphism (nsSNPs), found in an enrichment-based NGS study on HCM patients, against a whole-exome control and were able to explain 13–53% of HCM cases by studying four sarcomeric genes, which showed a significant excess of rare nsSNPs in the HCM cohort [55]. Van de Meerakker used linkage and haplotype analysis on PES of a DCM family in combination with conservation analyses and functional prediction before they could functionally verify an effect of the studied *TPMI* variant on the binding capacity to its actin partner [56]. These studies exemplify the potential NGS-based studies have to identify (new) disease variants (genes). However, such long-term studies do not have the potential to be of immediate benefit for patients. To ultimately translate all the findings into clinical care, physicians need a detailed report to judge the relevance of the identified known and novel variants in relation to the patient’s phenotype. A clearly structured more condensed version of the report should then be given to the patient to explain the diagnosis and subsequent therapy planning. A report should follow general principals of clinical genetic reporting and adhere to widely accepted guidelines for variant descriptions from the

Genome Variation Society (www.hgvs.org) [36]. The generation of such a report still requires a lot of manual work. Automated solutions like the knoSYS100 system from Knome (www.knome.com) have begun to emerge on the market. Similar “homegrown” solutions also already exist at some clinics, but those systems will have to evolve further to be regularly implemented. Currently, only genetic testing of genomic variants is performed. As mentioned above, data from the transcriptome and methylome can also be accessed quickly through NGS technology. In the future it might therefore be reasonable to integrate further molecular data into the course of diagnostics. If this is the case, the complexity will increase and reporting meaningful results will become even more difficult.

In summary, before a clinic decides on how to implement NGS into diagnostics, the following points should be considered. Will NGS be implemented at the local site or will it be used through service providers? If the latter is the case, how can data security of transferred genomic information be guaranteed? If a sequencing facility is installed at the local site, the required bioinformatics hardware has to grow with the increasing amount of data that will be produced [57]. This includes both computational power as well as storage capacity, which should be secure but also quickly accessible within the data analysis pipeline [58]. Enough long-term storage should be taken into account. Currently, the amount of NGS data produced is growing faster than computational power is expected to grow based on Moore’s law [59]. To close this gap it is possible to use cloud-based solutions from commercial vendors like the “Amazon Elastic Compute Cloud” (www.aws.amazon.com/ec2) as a computational resource. However, as mentioned above, data security as well as ethical aspects have to be considered before its implementation.

In conclusion, NGS-based diagnostics for cardiomyopathies and other inherited diseases are now technically feasible. A careful weighing of the points discussed will help to successfully implement such diagnostics in routine use in the near future to guide more personalized diagnosis and therapy planning.

References

- 1 Liu, L. *et al.* (2012) Comparison of next-generation sequencing systems. *Journal of Biomedicine & Biotechnology*, **2012**, 251364.
- 2 Mestan, K.K., Ilkhanoff, L., Mouli, S., and Lin, S. (2011) Genomic sequencing in clinical trials. *Journal of Translational Medicine*, **9**, 222.
- 3 Lupski, J.R. *et al.* (2010) Whole-genome sequencing in a patient with Charcot–Marie–Tooth neuropathy. *New England Journal of Medicine*, **362**, 1181–1191.
- 4 Ng, S.B. *et al.* (2010) Exome sequencing identifies the cause of a Mendelian disorder. *Nature Genetics*, **42**, 30–35.
- 5 Maron, B.J. *et al.* (2006) Contemporary definitions and classification of the cardiomyopathies: an American Heart Association Scientific Statement from the Council on Clinical Cardiology, Heart Failure and Transplantation Committee; Quality of Care and Outcomes Research and Functional Genomics and Translational Biology Interdisciplinary Working Groups; and Council on Epidemiology and Prevention. *Circulation*, **113**, 1807–1816.

- 6 Meder, B., Katus, H., and Rottbauer, W. (2009) Genetik der hypertrophischen Kardiomyopathie. *Journal für Kardiologie*, **16**, 274–278.
- 7 Haas, J., Katus, H.A., and Meder, B. (2011) Next-generation sequencing entering the clinical arena. *Molecular and Cellular Probes*, **25**, 206–211.
- 8 Watkins, H., Ashrafian, H., and Redwood, C. (2011) Inherited cardiomyopathies. *New England Journal of Medicine*, **364**, 1643–1656.
- 9 Geisterfer-Lowrance, A.A. *et al.* (1990) A molecular basis for familial hypertrophic cardiomyopathy: a beta cardiac myosin heavy chain gene missense mutation. *Cell*, **62**, 999–1006.
- 10 Wilde, A.A. and Behr, E.R. (2013) Genetic testing for inherited cardiac disease. *Nature Reviews Cardiology*, **10**, 571–583.
- 11 Perrot, A. *et al.* (2005) Prevalence of cardiac beta-myosin heavy chain gene mutations in patients with hypertrophic cardiomyopathy. *Journal of Molecular Medicine*, **83**, 468–477.
- 12 Kaski, J.P. *et al.* (2009) Prevalence of sarcomere protein gene mutations in preadolescent children with hypertrophic cardiomyopathy. *Circulation: Cardiovascular Genetics*, **2**, 436–441.
- 13 Shendure, J. *et al.* (2005) Accurate multiplex polony sequencing of an evolved bacterial genome. *Science*, **309**, 1728–1732.
- 14 Margulies, M. *et al.* (2005) Genome sequencing in microfabricated high-density picolitre reactors. *Nature*, **437**, 376–380.
- 15 Harris, T.D. *et al.* (2008) Single-molecule DNA sequencing of a viral genome. *Science*, **320**, 106–109.
- 16 Metzker, M.L. (2010) Sequencing technologies – the next generation. *Nature Reviews Genetics*, **11**, 31–46.
- 17 Quail, M.A. *et al.* (2012) A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers. *BMC Genomics*, **13**, 341.
- 18 Loman, N.J. *et al.* (2012) Performance comparison of benchtop high-throughput sequencing platforms. *Nature Biotechnology*, **30**, 434–439.
- 19 Albert, T.J. *et al.* (2007) Direct selection of human genomic loci by microarray hybridization. *Nature Methods*, **4**, 903–905.
- 20 Hodges, E. *et al.* (2007) Genome-wide *in situ* exon capture for selective resequencing. *Nature Genetics*, **39**, 1522–1527.
- 21 Okou, D.T. *et al.* (2007) Microarray-based genomic selection for high-throughput resequencing. *Nature Methods*, **4**, 907–909.
- 22 Porreca, G.J. *et al.* (2007) Multiplex amplification of large sets of human exons. *Nature Methods*, **4**, 931–936.
- 23 Stratton, M. (2008) Genome resequencing and genetic variation. *Nature Biotechnology*, **26**, 65–66.
- 24 Mamanova, L. *et al.* (2010) Target-enrichment strategies for next-generation sequencing. *Nature Methods*, **7**, 111–118.
- 25 Bodi, K. *et al.* (2013) Comparison of commercially available target enrichment methods for next-generation sequencing. *Journal of Biomolecular Techniques*, **24**, 73–86.
- 26 Shendure, J. and Lieberman Aiden, E. (2012) The expanding scope of DNA sequencing. *Nature Biotechnology*, **30**, 1084–1094.
- 27 Sboner, A., Mu, X.J., Greenbaum, D., Auerbach, R.K., and Gerstein, M.B. (2011) The real cost of sequencing: higher than you think! *Genome Biology*, **12**, 125.
- 28 Meder, B. *et al.* (2011) Targeted next-generation sequencing for the molecular genetic diagnostics of cardiomyopathies. *Circulation: Cardiovascular Genetics*, **4**, 110–122.
- 29 Mook, O.R. *et al.* (2013) Targeted sequence capture and GS-FLX Titanium sequencing of 23 hypertrophic and dilated cardiomyopathy genes: implementation into diagnostics. *Journal of Medical Genetics*, **50**, 614–626.
- 30 Dames, S., Durtschi, J., Geiersbach, K., Stephens, J., and Voelkerding, K.V. (2010) Comparison of the Illumina Genome Analyzer and Roche 454 GS FLX for resequencing of hypertrophic cardiomyopathy-associated genes. *Journal of Biomolecular Techniques*, **21**, 73–80.

- 31 Herman, D.S. *et al.* (2012) Truncations of titin causing dilated cardiomyopathy. *New England Journal of Medicine*, **366**, 619–628.
- 32 Ware, J.S. *et al.* (2013) Next generation diagnostics in inherited arrhythmia syndromes: a comparison of two approaches. *Journal of Cardiovascular Translational Research*, **6**, 94–103.
- 33 Haas, J., Barb, I., Katus, H.A., and Meder, B. (2014, in press) Targeted next-generation sequencing - the clinician's stethoscope for genetic disorders. *Biom. Med.*
- 34 Haas, J., Frese, K.S., Peil, B., Kloos, W., Keller, A., Nietsch, R., Feng, Z., Müller, S., Kayvanpour, E., Vogel, B., Sedaghat-Hamedani, F., Lim, W.K., Zhao, X., Fradkin, D., Köhler, D., Fischer, S., Franke, J., Marquart, S., Barb, I., Amr, A., Ehlermann, P., Mereles, D., Weis, T., Kremer, A., King, V., Wirsz, E., Isnard, R., Komajda, M., Serio, A., Grasso, M., Syrris, P., Wicks, E., Plagnol, V., Lopes, L., Gadgaard, T., Eiskjær, H., Jørgensen, M., Garcia-Giustiniani, D., Ortiz-Genga, M., Crespo-Leiro, M.G., Lekanne Dit Deprez, R.H., Christiaans, I., van Rijsingen, I.A., Wilde, A.A., Waldenstrom, A., Bolognesi, M., Bellazzi, R., Mörner, S., Lorenzo Bermejo, J., Monserrat, L., Villard, E., Mogensen, J., Pinto, Y.M., Charron, P., Elliott, P., Arbustini, E., Katus, H.A., and Meder, B. (2014, in press) Atlas of the clinical genetics of human dilated cardiomyopathy. *Eur Heart J*.
- 35 Sikkema-Raddatz, B. *et al.* (2013) Targeted next-generation sequencing can replace Sanger sequencing in clinical diagnostics. *Human Mutation*, **34**, 1035–1042.
- 36 Weiss, M.M. *et al.* (2013) Best practice guidelines for the use of next generation sequencing (NGS) applications in genome diagnostics: a national collaborative study of Dutch genome diagnostic laboratories. *Human Mutation*, **34**, 1313–1321.
- 37 Rehm, H.L. *et al.* (2013) ACMG clinical laboratory standards for next-generation sequencing. *Genetics in Medicine*, **15**, 733–747.
- 38 Scholtalbers, J. *et al.* (2013) Galaxy LIMS for next-generation sequencing. *Bioinformatics*, **29**, 1233–1234.
- 39 GenoLogis (2013) Clarity LIMS. GenoLogics Life Sciences Software, Victoria, BC.
- 40 Andreasen, C. *et al.* (2013) New population-based exome data are questioning the pathogenicity of previously cardiomyopathy-associated genetic variants. *European Journal of Human Genetics*, **21**, 918–928.
- 41 O'Rawe, J. *et al.* (2013) Low concordance of multiple variant-calling pipelines: practical implications for exome and genome sequencing. *Genome Medicine*, **5**, 28.
- 42 Adzhubei, I.A. *et al.* (2010) A method and server for predicting damaging missense mutations. *Nature Methods*, **7**, 248–249.
- 43 Ng, P.C. and Henikoff, S. (2003) SIFT: predicting amino acid changes that affect protein function. *Nucleic Acids Research*, **31**, 3812–3814.
- 44 Kumar, P., Henikoff, S., and Ng, P.C. (2009) Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nature Protocols*, **4**, 1073–1081.
- 45 Reva, B., Antipin, Y., and Sander, C. (2011) Predicting the functional impact of protein mutations: application to cancer genomics. *Nucleic Acids Research*, **39**, e118.
- 46 Gonzalez-Perez, A. and Lopez-Bigas, N. (2011) Improving the assessment of the outcome of nonsynonymous SNVs with a consensus deleteriousness score, Condel. *American Journal of Human Genetics*, **88**, 440–449.
- 47 Carter, H. *et al.* (2009) Cancer-specific high-throughput annotation of somatic mutations: computational prediction of driver missense mutations. *Cancer Research*, **69**, 6660–6667.
- 48 Choi, Y., Sims, G.E., Murphy, S., Miller, J.R., and Chan, A.P. (2012) Predicting the functional effect of amino acid substitutions and indels. *PLoS One*, **7**, e46688.
- 49 McLaren, W. *et al.* (2010) Deriving the consequences of genomic variants with the Ensembl API and SNP Effect Predictor. *Bioinformatics*, **26**, 2069–2070.

- 50 Dayem Ullah, A.Z., Lemoine, N.R., and Chelala, C. (2012) SNPnexus: a web server for functional annotation of novel and publicly known genetic variants (2012 update). *Nucleic Acids Research*, **40**, W65–W70.
- 51 Thusberg, J., Olatubosun, A., and Vihinen, M. (2011) Performance of mutation pathogenicity prediction methods on missense variants. *Human Mutation*, **32**, 358–368.
- 52 Thusberg, J. and Vihinen, M. (2009) Pathogenic or not? And if so, then how? Studying the effects of missense mutations using bioinformatics methods. *Human Mutation*, **30**, 703–714.
- 53 Castellana, S. and Mazza, T. (2013) Congruency in the prediction of pathogenic missense mutations: state-of-the-art web-based tools. *Briefings in Bioinformatics*, **14**, 448–459.
- 54 Meder, B. *et al.* (2009) A single serine in the carboxyl terminus of cardiac essential myosin light chain-1 controls cardiomyocyte contractility *in vivo*. *Circulation Research*, **104**, 650–659.
- 55 Lopes, L.R. *et al.* (2013) Genetic complexity in hypertrophic cardiomyopathy revealed by high-throughput sequencing. *Journal of Medical Genetics*, **50**, 228–239.
- 56 van deMeerakker, J.B. *et al.* (2013) A novel alpha-tropomyosin mutation associates with dilated and non-compaction cardiomyopathy and diminishes actin binding. *Biochimica et Biophysica Acta*, **1833**, 833–839.
- 57 Baker, M. (2010) Next-generation sequencing: adjusting to data overload. *Nature Methods*, **7**, 495–499.
- 58 Richter, B.G. and Sexton, D.P. (2009) Managing and analyzing next-generation sequence data. *PLoS Computational Biology*, **5**, e1000369.
- 59 Moore, G.E. (1965) Cramming more components onto integrated circuits. *Electronics*, **38**, 114–117.

