Chapter 7
Novel Next-Generation Sequencing Applications

Abstract Next-generation sequencing technologies have pushed the envelope beyond the primary goal of identifying the sequence of nucleotides within a given DNA molecule to a whole new multitude of applications. In this chapter, we describe select novel applications of next-generation sequencing in relation to large-scale sequencing-based projects, cell and cell compartments sequencing and disease-targeted sequencing.

7.1 Introduction

The applications of the next-generation sequencing technologies and the recently introduced third-generation sequencing methodologies are nearly limitless. The determination of the constituents of a DNA sequence itself was the primary aim of the first-generation of sequencing methods. With the availability of next-generation technologies, sequencing of the genome went from being a research aim to an important discovery tool. Thus, the utilization of whole genome sequencing (WGS), which is the primary application of these technologies, experienced a remarkable growth in the last few years. For instance, the number of genome sequencing projects in the Genome Online Database (GOLD) increased from 10,420 projects in May 2011 to 37,540 projects in January 2014 [1, 2]. The completed and published genomes in the above periods were 1,700 and 12,720 genomes, respectively, which demonstrated an incredible 720 % increase in a span of just 3 years. Clearly, this increase reflects the improved availability, affordability, and efficiency of the existing sequencers and methods.

The increased ease at which genome sequencing could be acquired opened the floodgates for applications and discoveries that went well beyond the initial goals of identifying the order of nucleotides or gene structure. Next-generation sequencing is presently being used in the WGS of humans [3], animals [4, 5], plants [6, 7], microbes [8, 9] and viruses [10]. In addition to WGS, next- and third-generation

S. El-Metwally et al., *Next Generation Sequencing Technologies and Challenges in Sequence Assembly*, SpringerBriefs in Systems Biology 7,
DOI 10.1007/978-1-4939-0715-1_7, © The Authors 2014

sequencing technologies are also employed in genome resequencing [11], RNA sequencing (RNA-seq) [12], whole exome sequencing (WES) [13], targeted sequencing [14], single-nucleotide variations discovery, analysis and validation [15], chromatin immunoprecipitation sequencing (ChIP-seq) [16], epigenetics [17], proteogenomics [18, 19], diseases and disorders targeted sequencing [20], mutations discovery [21], cancer research [22, 23] and numerous other clinical and health applications [24]. Several reports have extensively reviewed the genome sequencing applications [25, 26]. Furthermore, Nature Reviews Genetics has dedicated an ongoing article series to the applications of next-generation sequencing since 2009 [27]. Here, we will focus on the discussion of select novel applications that have been approached on a radically different scale since the advent of newer sequencing technologies.

7.2 Large-Scale Applications

Scientists design their research projects based on the availability and affordability of research tools and technologies. Thus, the availability of faster, cheaper, and more accurate tools and technologies leads to the planning of projects at an even higher level. The developments in genome sequencing technologies over the last decade have led to massive strides in sequencing power at an affordable cost and within a reasonable timeframe. This success has encouraged more expansive research projects where next-generation sequencing is used as a tool to discover diversities among individuals within large populations and to understand the fundamentals of life and biological systems. Here, we will take a few examples of large-scale genome projects that only became possible through the inception of next-generation sequencing and its subsequent development.

7.2.1 Genome 10K Project

In the year 2009, a group of genomics scientists established the Genome 10K Community of Scientists (G10KCOS) and announced the Genome 10K Project [28, 29]. The Genome 10K Project aims to sequence and annotate the genomes of about 10,000 vertebrate species that will amount to almost one species from each vertebrate genus. The project was inspired by the human genome project and the subsequent availability of 56 vertebrate (32 mammals and 24 nonmammalian) genomes that are appropriate for comparative genomic analyses [29]. The stated timeframe for the project is quite short as the community aims to assemble such a "genomic zoo" by 2015. The targeted species are distributed between all the vertebrates, including mammals, birds, non-avian reptiles, amphibians, and fishes. After 1 year, the G10KCOS announced the first 101 species to be sequenced [30]. Since fishes represent more than 50 % of extant vertebrates, the Genome 10K Project intends to sequence the genomes of about 4,000 fish species, 160 of which are currently in progress [31].

It is likely that the Genome 10K Project may take longer than expected. Nevertheless, it is a staggering effort that promises unique comparative study opportunities that is only possible through the use of modern genome sequencing technologies.

7.2.2 Tree of Life Sequencing Project

Another example of large-scale genome sequencing projects is the Tree of Life Sequencing Project that was announced by Beijing Genome Institute (BGI) in 2010. BGI has the most powerful sequencing capacity worldwide, and is the main contributor to the 1000 Genomes and Genome 10K projects as well. The Tree of Life Sequencing Project is also known as the 1000 Plant & Animal Reference Genomes Project, a name that is more descriptive of the intended goals of the venture. The project aims to target 1,000 reference genomes from 500 animals and 500 plants of various economically and scientifically important species such as rice, silkworm, cucumber, panda, camel, oyster, ant, grouper, goose, crested ibis, and potato genomes. To date, 106 genomes have been completed and published while another 200 are in progress, representing about 30 % of the targeted species [32].

7.3 Cell and Cell Compartments Applications

The projects discussed in the previous section shared the tendency to sequence a huge number of organisms and provide their genomes as reference genomes. In contrast, we will now examine the application of next-generation sequencing on a much smaller scale, such as a single cell or even a cell compartment. The main aims of such applications are to sequence the genomes of species that are difficult to grow in the lab environment, or when the availability of samples is limited. Another interesting possibility is the determination of the heterogeneity between single cells in normal or tumorous tissues.

7.3.1 Single-Cell Genome Sequencing

Preparation of sequencing samples is initiated with a group of cells, e.g., cell cultures of bacteria or archaea. However, culturing attempts have failed in the case of several microorganisms, making full genome sequencing of such organisms unlikely [33, 34]. Thus, methods to sequence a single cell were developed using PCR-based amplification of the single bacterial cell genome with accuracy approaching 97 % [35]. Another technique that increased accuracy to 99.6 % [36] involved PCR-based amplification with multiple displacement amplification (MDA) [34] followed by post-amplification normalization and assembly with the reference genome. These methods can be used to sequence the genomes of either single cells or individual cells from a variety of samples (with different treatments or from different environmental sources).

Single-cell sequencing applications have also expanded to include the study of diseases with genetic alterations and to find variations (heterogeneity) between different cells in diseased tissue. For instance, the amount of cancer-related genomic mutations in the Catalogue Of Somatic Mutations In Cancer (COSMIC) database number over one million to date [37]. The heterogeneity of tumor cells can result in several complications such as developing rare chemo-resistant cells that can resist chemotherapy. Such cells can regrow and result in the formation of a chemo-resistant tumor [38]. Several attempts have been made to apply single-cell sequencing to cancer genomics, allowing the possibility to sequence up to 200 single cells independently during a single run [39]. The numerous single-cell sequencing applications in cancer can include the pinpointing of chemo-resistant cells, the early detection of tumor cells, measuring intratumor heterogeneity, monitoring of circulating tumor cells (CTCs) and in drug target discovery [40, 41]. Furthermore, the techniques may also be utilized to develop a guided form of chemotherapy that is appropriate against the measured heterogeneity of the tumor [39]. In the later sections, we will discuss further details on the applications of sequencing in cancer.

7.3.2 Mitochondrial Genome Sequencing

Mitochondria are cellular organelles that can be found in eukaryotic cells. They are responsible for producing most of the cell's energy by supplying it with adenosine triphosphate (ATP) through the phosphorylation of adenosine diphosphate (ADP). Mitochondria have their own genome and genetics that are independent from the cell nucleus genome. Therefore, it has its own proteome that is about 615 proteins [41]. Most of the mitochondria are inherited from the mother, and there is group of diseases known as mitochondrial diseases that are caused by dysfunctional mitochondria or genes that are inherited through the mitochondrial genome [42, 43]. These structures are also attributed to play an important role in aging and cancer [44, 45]. Moreover, they have a special genetic code for tryptophan and methionine as well as a distinct stop codon. This allows the mitochondrial genome to be perfectly suited for forensic investigations and human phylogenic studies [44, 45]. Hence, advancements in next-generation sequencing [43, 46] have been aptly reflected in the utility of human mitochondrial genome sequencing during forensic investigations and cancer [45, 46] as well as the study of plants [47] and fish [48].

7.4 Disease-Targeted Sequencing

Several diseases are associated with genetic mutations or genetic disorders while others are inherited from carrier parents to their offspring. The ongoing discovery of disease-related genes has made disease-targeted sequencing tests an important diagnostic tool [49]. With Sanger sequencing, tests were designed for diseases with a

Table 7.1 Clinically available disease-targeted tests[a]

Disease area	Disease type	Number of genes
Cancer	Hereditary cancers (for example, breast, colon, and ovarian)	10–50
Cardiac diseases	Cardiomyopathies	50–70
	Arrhythmias (for example, long QT syndrome)	10–30
	Aortopathies (for example, Marfan's syndrome)	10
Immune disorders	Severe combined immunodeficiency syndrome	18
	Periodic fever	7
Neurological, neuromuscular and metabolic disorders	Ataxia	40
	Cellular energetics, metabolism	656
	Congenital disorders of glycosylation	23–28
	Dementia (for example, Parkinson's disease and Alzheimer's disease)	32
	Developmental delay, autism, intellectual disability	30–150
	Epilepsy	53–130
	Hereditary neuropathy	34
	Microcephaly	11
	Mitochondrial disorders	37–450
	Muscular dystrophy	12–45
Sensory disorders	Eye disease (for example, retinitis pigmentosa)	66–140
	Hearing loss and related syndromes	23–72
Other	Rasopathies (for example, Noonan's syndrome)	10
	Pulmonary disorders (for example, cystic fibrosis)	12–40
	Short stature	12

[a] Data is derived from [49]

single causative gene in order to confirm the diagnosis. On the other hand, designing tests for diseases with enormous genetic heterogeneity is far more difficult [49]. With the introduction of next-generation sequencing, the significant increase in throughput and reduction in technical costs greatly aid the design of tests for a wide spectrum of diseases and genetic disorders, as well as the discovery of new disease-related genes and mutations (Table 7.1). In this section, we will introduce some of the recent applications of next-generation sequencing in understanding inherited and complex diseases, including the study of disease-related genes and mutations.

7.4.1 Sequencing in Cancer

Cancer is widely known to be associated with somatic mutations [22]. The Sanger Institute launched the Cancer Genome Project (CGP) as one of the earliest attempts to identify cancer genes and mutations [50]. The CGP currently represents one of the main resources of cancer genomics and mutations with its several databases and resources, including the COSMIC database [51], the Cancer Gene Census [52], COSMIC whole genomes and the COSMIC cell-line project [37]. To date, over a

million identified mutations in cancer have been cataloged in the COSMIC database, including all types of known genetic mutations such as single-nucleotide mutations, insertions, deletions, and chromosomal rearrangements [37, 51]. Although the primary technology utilized at the commencement of the CGP was Sanger sequencing, the project also utilized the power of next-generation sequencing in later phases.

Several other large-scale projects have been conceived through international consortiums aided by public and private funding. These projects also aim to identify cancer-related mutations and genes as well as categorize findings based on importance and recurrence. For example, the International Cancer Genome Consortium (ICGC) is a huge publicly funded cancer genome-sequencing project. The ICGC aims to sequence the whole genome of 50 different types and subtypes of cancer that are clinically important [53]. The most recent data release from the project (Release 14) provides the results of 41 different cancer projects from over 8,500 donors. In this case, sequencing studies resulted in the identification of over two million mutations from 54,682 mutated genes.

With relation to privately funded projects, the Pediatric Cancer Genome Project (PCGP) was announced in 2010 by St. Jude Children's Research Hospital and the Genome Institute at Washington University [54]. This project targeted the sequencing of 600 pediatric tumors and matched non-tumor germline samples (totaling 1,600 genomes) with high resolution sequencing in an aim to catalog somatic mutations of pediatric tumors and define the major subtypes in pediatric cancers [54]. The most recent data release from the PCGP (June 2013) contained the whole genomes of 15 different cancer types from over 360 patients that were analyzed and revealed novel findings [55].

7.4.2 Sequencing in Inherited Human Diseases

Inherited human diseases are disorders that result from single-gene mutations. They are also known as monogenic disorders or Mendelian disorders. There are around 5,000 known monogenic disorders though the genetic causes of most of them are still unknown [56]. Most of these cases resulted from exonic mutations (mutations that occur in the exon) or splice-site mutations (mutations that affect the splicing pattern of the mRNA). Both types of mutations affect the resulting protein sequence following translation of the affected gene [57]. Thus, whole exome sequencing (WES) using next-generation sequencing is an efficient methodology to identify both these types of mutations without the need of whole genome sequencing (WGS). Furthermore, the utilization of WES saves time and reduces cost since the human exome represents ~1 % of the human genome. However, certain other mutations that cannot be identified without sequencing the whole genome may also result from deletions [57]. The 1,000 Mendelian Disorders Project is a large-scale effort at the Beijing Genome Institute (BGI) that aims to sequence the genome of 1,000 Mendelian disorders to identify the causative genes behind them using next-generation sequencing rather than traditional techniques such as positional cloning, physical mapping, and candidate-gene sequencing [56].

7.4.3 Sequencing in Complex Human Diseases

Next-generation sequencing has provided novel approaches in locating common and rare variants that influence the risk of developing complex diseases such as cancer, diabetes, cardiovascular disease, and psychiatric disorders [25]. Several Genome-Wide Association Studies (GWAS) have used next-generation sequencing technologies in examining complex trait genetics [58, 59]. Such studies demonstrated the utility of next-generation sequencing applications in understanding complex diseases such as hypertrophic cardiomyopathy [59], brain disease [60] and diabetes [61]. Moreover, the investigations provided novel insight into understanding the genetics mechanisms behind disorders of sex development (DSD) [62].

References

1. Kyrpides NC (1999) Genomes OnLine Database (GOLD 1.0): a monitor of complete and ongoing genome projects world-wide. Bioinformatics 15 (9):773-774. doi:btc112
2. Pagani I, Liolios K, Jansson J, Chen IM, Smirnova T et al. (2012) The Genomes OnLine Database (GOLD) v.4: status of genomic and metagenomic projects and their associated metadata. Nucleic Acids Res 40 (Database issue):D571-579. doi:10.1093/nar/gkr1100
3. Yamey G (2000) Scientists unveil first draft of human genome. BMJ 321 (7252):7
4. Michelizzi VN, Dodson MV, Pan Z, Amaral ME, Michal JJ et al. (2010) Water buffalo genome science comes of age. Int J Biol Sci 6 (4):333-349
5. Edwards CJ, Magee DA, Park SD, McGettigan PA, Lohan AJ et al. (2010) A complete mitochondrial genome sequence from a mesolithic wild aurochs (Bos primigenius). PLoS One 5 (2):e9255. doi:10.1371/journal.pone.0009255
6. Cao J, Schneeberger K, Ossowski S, Gunther T, Bender S et al. (2011) Whole-genome sequencing of multiple Arabidopsis thaliana populations. Nat Genet 43 (10):956-963. doi:10.1038/ng.911
7. Matsumoto T, Wu J, Antonio BA, Sasaki T (2008) Development in rice genome research based on accurate genome sequence. Int J Plant Genomics 2008:348621. doi:10.1155/2008/348621
8. Shendure J, Porreca GJ, Reppas NB, Lin X, McCutcheon JP et al. (2005) Accurate multiplex polony sequencing of an evolved bacterial genome. Science 309 (5741):1728-1732. doi:1117389
9. Otero JM, Vongsangnak W, Asadollahi MA, Olivares-Hernandes R, Maury J et al. (2010) Whole genome sequencing of Saccharomyces cerevisiae: from genotype to phenotype for improved metabolic engineering applications. BMC Genomics 11:723. doi:10.1186/1471-2164-11-723
10. Ren X, Yang F, Hu Y, Zhang T, Liu L et al. (2013) Full genome of influenza A (H7N9) virus derived by direct sequencing without culture. Emerg Infect Dis 19 (11):1881-1884. doi:10.3201/eid1911.130664
11. Costa GG, Cardoso KC, Del Bem LE, Lima AC, Cunha MA et al. (2010) Transcriptome analysis of the oil-rich seed of the bioenergy crop Jatropha curcas L. BMC Genomics 11:462. doi:10.1186/1471-2164-11-462
12. Wang H, Zou Z, Wang S, Gong M (2013) Global Analysis of Transcriptome Responses and Gene Expression Profiles to Cold Stress of Jatropha curcas L. PLoS One 8 (12):e82817. doi:10.1371/journal.pone.0082817
13. Lai CC, Yeh YH, Hsieh WP, Kuo CT, Wang WC et al. (2013) Whole-Exome Sequencing to Identify a Novel LMNA Gene Mutation Associated with Inherited Cardiac Conduction Disease. PLoS One 8 (12):e83322. doi:10.1371/journal.pone.0083322
14. Bazak L, Haviv A, Barak M, Jacob-Hirsch J, Deng P et al. (2013) A-to-I RNA editing occurs at over a hundred million genomic sites, located in a majority of human genes. Genome Res. doi:gr.164749.113

15. Rellstab C, Zoller S, Tedder A, Gugerli F, Fischer MC (2013) Validation of SNP allele frequencies determined by pooled next-generation sequencing in natural populations of a non-model plant species. PLoS One 8 (11):e80422. doi:10.1371/journal.pone.0080422
16. Marinov GK, Kundaje A, Park PJ, Wold BJ (2013) Large-Scale Quality Analysis of Published ChIP-seq Data. G3 (Bethesda). doi:g3.113.008680v1
17. Milavetz B, Kallestad L, Woods E, Christensen K, Gefroh A et al. (2013) Erratum: Transcription and replication result in distinct epigenetic marks following repression of early gene expression. Front Genet 4:259. doi:10.3389/fgene.2013.00259
18. Helmy M, Sugiyama N, Tomita M, Ishihama Y (2012) Mass spectrum sequential subtraction speeds up searching large peptide MS/MS spectra datasets against large nucleotide databases for proteogenomics. Genes Cells 17 (8):633-644. doi:10.1111/j.1365-2443.2012.01615.x
19. Helmy M, Tomita M, Ishihama Y (2012) Peptide identification by searching large-scale tandem mass spectra against large databases: bioinformatics methods in proteogenomics. Genes Genome Genomics 6:76-85
20. Clement JA, Toulza E, Gautier M, Parrinello H, Roquis D et al. (2013) Private Selective Sweeps Identified from Next-Generation Pool-Sequencing Reveal Convergent Pathways under Selection in Two Inbred Schistosoma mansoni Strains. PLoS Negl Trop Dis 7 (12):e2591. doi:10.1371/journal.pntd.0002591
21. Damerla RR, Chatterjee B, Li Y, Francis RJ, Fatakia SN et al. (2013) Ion Torrent sequencing for conducting genome-wide scans for mutation mapping analysis. Mamm Genome. doi:10.1007/s00335-013-9494-7
22. Helmy M, Sugiyama N, Tomita M, Ishihama Y (2010) Onco-proteogenomics: a novel approach to identify cancer-specific mutations combining proteomics and transcriptome deep sequencing. Genome Biol 11. doi:10.1186/Gb-2010-11-S1-P17
23. Patel L, Parker B, Yang D, Zhang W (2013) Translational genomics in cancer research: converting profiles into personalized cancer medicine. Cancer Biol Med 10 (4):214-220. doi:10.7497/j.issn.2095-3941.2013.04.005
24. Chang F, Li MM (2013) Clinical application of amplicon-based next-generation sequencing in cancer. Cancer Genet. doi:S2210-7762(13)00142-7
25. Pareek CS, Smoczynski R, Tretyn A (2011) Sequencing technologies and genome sequencing. J Appl Genet 52 (4):413-435. doi:10.1007/s13353-011-0057-x
26. Soon WW, Hariharan M, Snyder MP (2013) High-throughput sequencing for biology and medicine. Mol Syst Biol 9:640. doi:10.1038/msb.2012.61
27. Nature Reviews Genetics Article Serie (2009) Applications of next-generation sequencing. http://www.nature.com/nrg/series/nextgeneration/index.html. Accessed 10-01-2014
28. Genome 10K Project (2009) Genome 10K Project Home Page. https://genome10k.soe.ucsc.edu/. Accessed 10-01-2014
29. Genome 10K Scientists (2009) Genome 10K: a proposal to obtain whole-genome sequence for 10,000 vertebrate species. J Hered 100 (6):659-674. doi:10.1093/jhered/esp086
30. Wagman B (2010) Genome 10K project announces first 101 species for genome sequencing. http://cbse.soe.ucsc.edu/news/article/1820?ID=1820. Accessed 10-01-2014
31. Bernardi G, Wiley EO, Mansour H, Miller MR, Orti G et al. (2012) The fishes of Genome 10K. Mar Genomics 7:3-6. doi:10.1016/j.margen.2012.02.002
32. Li Q, Li N, Hu X, Li J, Du Z et al. (2011) Genome-wide mapping of DNA methylation in chicken. PLoS One 6 (5):e19428. doi:10.1371/journal.pone.0019428
33. Lasken RS (2012) Genomic sequencing of uncultured microorganisms from single cells. Nat Rev Microbiol 10 (9):631-640. doi:10.1038/nrmicro2857
34. Dean FB, Hosono S, Fang L, Wu X, Faruqi AF et al. (2002) Comprehensive human genome amplification using multiple displacement amplification. Proc Natl Acad Sci U S A 99 (8):5261-5266. doi:10.1073/pnas.082089499
35. Woyke T, Tighe D, Mavromatis K, Clum A, Copeland A et al. (2010) One bacterial cell, one complete genome. PLoS One 5 (4):e10314. doi:10.1371/journal.pone.0010314
36. Rodrigue S, Malmstrom RR, Berlin AM, Birren BW, Henn MR et al. (2009) Whole genome amplification and de novo assembly of single bacterial cells. PLoS One 4 (9):e6864. doi:10.1371/journal.pone.0006864

37. Forbes SA, Bindal N, Bamford S, Cole C, Kok CY et al. (2011) COSMIC: mining complete cancer genomes in the Catalogue of Somatic Mutations in Cancer. Nucleic Acids Res 39 (Database issue):D945-950. doi:10.1093/nar/gkq929
38. Oakman C, Santarpia L, Di Leo A (2010) Breast cancer assessment tools and optimizing adjuvant therapy. Nat Rev Clin Oncol 7 (12):725-732. doi:10.1038/nrclinonc.2010.170
39. Navin N, Hicks J (2011) Future medical applications of single-cell sequencing in cancer. Genome Med 3 (5):31. doi:10.1186/gm247
40. Spaethling JM, Eberwine JH (2013) Single-cell transcriptomics for drug target discovery. Curr Opin Pharmacol 13 (5):786-790. doi:10.1016/j.coph.2013.04.011
41. Campbell NA (ed) (2003) Biology: Exploring Life. Pearson Prentice Hall, Upper Saddle River, NJ
42. Dimauro S, Davidzon G (2005) Mitochondrial DNA and disease. Ann Med 37 (3):222-232. doi:R53R2R64844U5876
43. Lyons EA, Scheible MK, Sturk-Andreaggi K, Irwin JA, Just RS (2013) A high-throughput Sanger strategy for human mitochondrial genome sequencing. BMC Genomics 14:881. doi:10.1186/1471-2164-14-881
44. Wallace DC (2010) Mitochondrial DNA mutations in disease and aging. Environ Mol Mutagen 51 (5):440-450. doi:10.1002/em.20586
45. Brandon M, Baldi P, Wallace DC (2006) Mitochondrial mutations in cancer. Oncogene 25 (34):4647-4662. doi:1209607
46. Parson W, Strobl C, Huber G, Zimmermann B, Gomes SM et al. (2013) Reprint of: Evaluation of next generation mtGenome sequencing using the Ion Torrent Personal Genome Machine (PGM). Forensic Sci Int Genet 7 (6):632-639. doi:10.1016/j.fsigen.2013.09.007
47. Fajardo D, Schlautman B, Steffan S, Polashock J, Vorsa N et al. (2013) The American cranberry mitochondrial genome reveals the presence of selenocysteine (tRNA-Sec and SECIS) insertion machinery in land plants. Gene. doi:S0378-1119(13)01648-X
48. Hester J, Atwater K, Bernard A, Francis M, Shivji MS (2013) The complete mitochondrial genome of the basking shark Cetorhinus maximus (Chondrichthyes, Cetorhinidae). Mitochondrial DNA. doi:10.3109/19401736.2013.845762
49. Rehm HL (2013) Disease-targeted sequencing: a cornerstone in the clinic. Nat Rev Genet 14 (4):295-300. doi:10.1038/nrg3463
50. Davies H, Bignell GR, Cox C, Stephens P, Edkins S et al. (2002) Mutations of the BRAF gene in human cancer. Nature 417 (6892):949-954. doi:10.1038/nature00766
51. Alexandrov LB, Nik-Zainal S, Wedge DC, Aparicio SA, Behjati S et al. (2013) Signatures of mutational processes in human cancer. Nature 500 (7463):415-421. doi:10.1038/nature12477
52. Futreal PA, Coin L, Marshall M, Down T, Hubbard T et al. (2004) A census of human cancer genes. Nat Rev Cancer 4 (3):177-183. doi:10.1038/nrc1299
53. Hudson TJ, Anderson W, Artez A, Barker AD, Bell C et al. (2010) International network of cancer genome projects. Nature 464 (7291):993-998. doi:10.1038/nature08987
54. Chen X, Stewart E, Shelat AA, Qu C, Bahrami A et al. (2013) Targeting oxidative stress in embryonal rhabdomyosarcoma. Cancer Cell 24 (6):710-724. doi:10.1016/j.ccr.2013.11.002
55. Zhang J, Wu G, Miller CP, Tatevossian RG, Dalton JD et al. (2013) Whole-genome sequencing identifies genetic alterations in pediatric low-grade gliomas. Nat Genet 45 (6):602-612. doi:10.1038/ng.2611
56. Liu Y, Gao M, Lv YM, Yang X, Ren YQ et al. (2011) Confirmation by exome sequencing of the pathogenic role of NCSTN mutations in acne inversa (hidradenitis suppurativa). J Invest Dermatol 131 (7):1570-1572. doi:10.1038/jid.2011.62
57. Kuhlenbaumer G, Hullmann J, Appenzeller S (2011) Novel genomic techniques open new avenues in the analysis of monogenic disorders. Hum Mutat 32 (2):144-151. doi:10.1002/humu.21400
58. Day-Williams AG, Zeggini E (2011) The effect of next-generation sequencing technology on complex trait research. Eur J Clin Invest 41 (5):561-567. doi:10.1111/j.1365-2362.2010.02437.x
59. Voelkerding KV, Dames S, Durtschi JD (2010) Next generation sequencing for clinical diagnostics-principles and application to targeted resequencing for hypertrophic cardiomyopathy: a paper from the 2009 William Beaumont Hospital Symposium on Molecular Pathology. J Mol Diagn 12 (5):539-551. doi:10.2353/jmoldx.2010.100043

60. Zoghbi HY, Warren ST (2010) Neurogenetics: advancing the "next-generation" of brain research. Neuron 68 (2):165-173. doi:10.1016/j.neuron.2010.10.015
61. Nejentsev S, Walker N, Riches D, Egholm M, Todd JA (2009) Rare variants of IFIH1, a gene implicated in antiviral responses, protect against type 1 diabetes. Science 324 (5925):387-389. doi:10.1126/science.1167728
62. Bashamboo A, Ledig S, Wieacker P, Achermann JC, McElreavey K (2010) New technologies for the identification of novel genetic markers of disorders of sex development (DSD). Sex Dev 4 (4-5):213-224. doi:10.1159/000314917