Research Article   **Open Access**

High Throughput Sequencing Advances and Future Challenges

Tariq Pervaiz[1,2], Abolfazl Lotfi[1], Muhammad Salman Haider[1], Jia Haifang[1] and Jinggui Fang[1*]

[1]Key Laboratory of Genetics and Fruit Development, College of Horticulture, Nanjing Agricultural University, Nanjing, China

[2]Department of Agriculture and Food Tech. Karakorum International University, Gilgit Pakistan

*Corresponding author: Jinggui Fang, Key Laboratory of Genetics and Fruit Development, College of Horticulture, Nanjing Agricultural University, Nanjing, China, Tel: +86 25 8439 5366; E-mail: fanggg@njau.edu.cn

Abstract

High throughput sequencing (HTS) technologies were developed into indispensable for genomic investigation and recent hottest topic for research in the field of genomics, which can generate over 100 times more data in comparison with the most complicated capillary sequencers. Recent advances and developments in HTS using next generation sequencing techniques have become essential in the studies of digital gene expression profiling, in epigenomics, genomics, and transcriptomics. These methodologies are dexterous of sequencing multiple DNA molecules in corresponding; facilitate hundreds of millions of DNA molecules to be sequenced within a short period of time. Though, the expenses and time period have been significantly reduced; the inaccurate profiles and boundaries of the new policy differ considerably from those of earlier reported sequencing techniques. The technical developments and decreasing cost of NGS (Next Generation Sequencing) technology have made RNA sequencing (RNA-seq) as a worldwide popular technique for gene expression projects. Various approaches have been done for the standardization of RNA sequencing data, which have been materialized in the reports, contradictory, both in the type of bias modification and in the statistical approach. On the other hand, as data persistently build up, there has been no apparent consensus on the proper normalization techniques to be used or the impact of chosen methods on the downstream analysis. In the present article, we mentioned the key features of HT-NGS like, Key HTS platforms and different sequencing applications, ethical limitation and future prospective.

Keywords: Next generation sequences; High throughput; RNA sequences; Genomics; miRanalyzer

Background

The advancement of high-throughput sequencing (HTS) techniques remarkable advances in technologies have fundamentally changed to understand the genetic and epigenetic molecular bases underlying human health and plant diseases [1-3]. The influence of these methodologies was developed from the constant sequencing of genomic regions of interest, for example, exons and protein binding sites [4,5]. The methodologies involve dispensation large number of sequencing reads controlled in unprocessed data sets ranging from a large number of megabytes to over 25 gigabytes [6]. HTS is a novel and rapid advanced sequencing technology which is commonly used in transcriptomics, genomics and epigenomics [7], most recent advancements and approaches during in the last decades, that alternative sequencing have been develop into available strategies [8,9] which emphasis to entirely redefine "High-throughput Sequencing Technology." The present technological approaches do better than the previously reported Sanger-Sequencing tools with a factor of addition of 100–1000 in day by day throughput, and ultimately, by lowering expenses of one million (1 Mb) nucleotides to 4-0.1% of that making sequencing technologies mainstream supplementary through Sanger sequencing [10,11]. To reproduce massive changes, many researchers, recent reviews and companies apply the terminology "NGS" (next-generation sequencing) as an alternative of high-throughput sequencing (HTS) [12]. Current developments in sequencing techniques have been dramatically changed the field of genomics during the present technological development, making it promising even particular research members to generate gigantic amount of sequence data very fast at a considerably subsidized budget and expenses. These HTS techniques make full genome sequencing, transcript quantification, resequencing and deep transcriptome sequencing are available for future researcher.

MicroRNAs (miRNAs) play essential regulatory functions in several organisms through direct cleavage of transcripts, chromatin modification or translational repression and modulate gene expression in both plants and animals. Identification of miRNAs has been carried out in most plant species [13]. Since the first detection of plant miRNAs from Arabidopsis, plant miRNAs have been extremely investigated using experimental and computational techniques [14]. According to miRBase release, the large number of miRNAs were identified and reached up to 4011 [15]. Due to the significant role of the miRNAs in plant natural processes, ranging from organ differentiation to biological and environmental stress responses [16]. The recent advances of computational, analytical and experimental advancements, attention in these undersized molecules has been considerably improved during last and resent decades. Up to date, 25141 mature miRNA sequences from 193 various species (viruses to human) reported in the miRBase catalog. The Plant MicroRNA repository PMRD, released on June 11, 2012, having 10597 miRNAs discovered in 127 plant species. Currently HTS with NGS tools becomes precious in digital gene expression profiling [17]. Many researchers to recognize and verify the expression of new and conserved tissue or developmental stage-specific levels of miRNAs [18] have used these tools. In addition, methylation sequencing and chromatin immune precipitation sequencing might be helpful to categorize epigenetic changes, while ribosome sequencing used to resolve mRNA transcripts are dynamically individual translated.

Especially, woody plants with high heterozygosity, like in Chinese bayberry, entire genome sequencing involved long-term and costly; as a result it is presently limited to only some species. As an alternative, it has been more helpful to get information of UniGenes throughout transcriptome sequencing [19,20]. Compared with conventional lab techniques, RNA-Seq is a HTS technology, overcoming the short comes of microarrays tools, in exploring unidentified genes. Moreover, thus having recompense in investigating transcriptome fine structures, for instance, splice junction variation and detection of allele-specific expression [21]. In this review, we summarized the useful high throughput sequencing technologies, their application, limitation and also the future of HTS.

Key HTS Platforms and Applications

Sanger capillary sequencing: Contemporary SCS extensively used in the GE Healthcare Mega BACE or Applied Biosystems 3xxx series instrument are supported by the general similar system were used during 1977 for the genome (wX174) [22]. Initially, millions of duplicates of the structure to be indomitable are amplified or purified, contingent on the sequence basis. Reverse strand synthesis is presented on these duplicates using an identified priming sequence upstream to be indomitable and a combination of dideoxy-nucleotides (ddNTP, modified nucleotides missing a hydroxylgroup in sugar at the third carbon atom) and deoxy-nucleotides. The non-reversible execution of the extension reaction, dNTP/ddNTP mixture causes random producing from the multiple copies of molecules extended to varied size. Subsequently denaturation and cleanup of free moving enzyme, primers and the nucleotides, while the resulting molecules are arranged by their molecular weight (analogous of the point of termination) and the markers involved to the terminating ddNTPs is read out successively in order formed by the categorization step. With the application of existing Sanger sequencing techniques, it is precisely capable of 384 sequences [23] ranging from 600 to 1,000 nt, in length [24,25] though, the present 384-capillary systems are exceptional. The additional ordinary 96-capillary apparatus produce a utmost of more or less 6 Mb of DNA sequence apiece date, with expenses for consumables cost about $500/1 Mb.

De novo Sequencing

De novo sequencing is applied for generation of the DNA sequence from DNA molecule with no any aforementioned knowledge regarding the sequences. For genome research works, enormously high throughput level and high-end robotics are mandatory in order to provide the sequencing workflow. During the last decade next-generation sequencing technologies considerably emphasized on "Moore's Law", which predicted that the high throughput of DNA sequencing in faster rate than the development rate of computer technology, as researchers face difficulties to smooth load and operate in computer memory system. There is demand for the de novo assemblers to capably hold the large scale of sequencing data using scalable commodity servers in the clouds. Chang, Chen [26], reported Cloud Brush, analogous algorithm that runs on the Map Reduce framework of cloud computing high-throughput sequencing data for de novo assembly. The algorithms apply Myers's bi-directed string graphs as its basis and consist of two major phases: graph simplification and construction. De novo Sequencing applications in various workflow strategies were developed to perform de novo sequencing. They include Primer walking, Shotgun sequencing, using

transposons to randomly prime sites for sequencing, PCR amplification of template, nested deletions and mRNA sequencing.

De novo complete or partial genome sequencing might be address through a variety of common approaches. Initial generation of genomic sequence of new species and detailed genetic analysis were only possible after de novo sequencing has been performed, RNA-Seq is recently developed as an influential, high-throughput sequencing technology that used to produce millions of short sequence reads in short time period by deep sequencing. Long reads are more useful and paired end reads are essential, which allowing gene expression profiling that divulge several new transcribed sections; splice forms, SNPs and accurate localization of transcription boundaries. ESTs are incomplete sequences retrieved from complementary DNA (cDNA). A number of ESTs might be produced from a single gene and represent gene expression in all individual samples [27]. The whole transcription units representing by Full-length cDNAs are more efficient than incomplete sequences for genome annotation and transcriptome analysis [28,29]. Full-length cDNAs chosen and can be developed stand on the 59-cap, a unique traits of mRNA structure [30,31]. Additionally, the genes expected from the de-novo assemblies have to be authenticated to guarantee the efficiency of the assembly process. For the reason that reverse transcriptase PCR (RT-PCR) assist the detection and quantification of target mRNA transcripts, Kim, Lim [32] applied RT-PCR techniques to predicted tissue-specific candidate genes in order to confirm the consistency of transcriptome assembly in Brassica oleracea. RT-PCR used to explore the tissue-specific genes discovered by de novo assembly and analysis of deep-sequencing data could be indicating to experimentally certify the continuation of the assembled genes. Parallel to adornment the low-quality bases at the end of reads, amalgamation of the contigs produced by multiple assemblies could also improve the assembly outcome [33,34]. Tissue-specific genes are especially practical and expressed in specific cell types or tissues. Not only experimental validation of de novo assembled genes, but similarly. Thus, the knowledge directs us to altitudinal or time-course expression patterns show when and where particular genes are working. Thus, the evidence allows us to assume link between chronological or growth stage-specific manifestation, genes and tissues, and innovative gene functions [32].

Metagenomics

Metagenomics termed as MPS (Massive Parallel Sequencing) of metagenome DNA exclusive of targeted intensification, facilitate a significant increase in the capacity of data created. Farther more, the expenditures of MPS is declining quickly. Untargeted MPS used extracting entire RNA or DNA from the tissue or population of interest. cDNA or DNA is afterward sequenced, exclusive of besieged amplification, using a particularly parallel pyro-sequencing podium, for instance the Illumina GAIIx. Currently, untargeted Massive Parallel Sequencing of rumen microbial populace have been discovered various innovative gene sequences used for deep sequencing of single pooled sample [35]. Conversely, discrepancies among individuals of the similar category have not yet been revealed due to in rumen discrepancy (sampling error) or technical or factual biological variation. The method developed for evaluation the dissimilarity in rumen meta-genome contours involves a "reference meta-genome", for instance a succession of contig sequences starting from previously reported experimentations, in which sample sequence reads were aligned. The "rumen meta-genome profile" is then the total of the reads that make parallel to every contig. For instance, in the database

200,000 contigs, the profile vector of counts will be a 200,000 × 1. These contours can then be analyzed through hierarchical clustering and bootstrap analysis or linear mixed models [36].

Whole Genome/Targeted Re-sequencing

Targeted genome improvement is influential means for assembly of the substantial throughput of new DNA-sequencing apparatus. In present, a scalable and simple procedure for multiplex enlargement of target regions built on the selector methods. The modernized version exhibits enhanced exposure and compatibility with NGS library-construction method for shotgun sequencing through NGS podiums [37].

NGS techniques, which facilitate the fast production of whole genome sequences have modernized genetic research [38]. By means of the materialization of economical work surface of NGS platforms, genomic investigation are at present being carried out in applied or translational study labs as a substitute of state-of-the art genome center [39,40]. Although these technological progress, comprehensive finishing and de novo sequence (genome closure) assembly persist to urge researchers [41]. Therefore, outsized incomplete genome sequences have been launched to databanks [42]. Rationalized techniques to assemble high-quality whole genome sequences were needed, particularly in the microbial genomics of Gram-negative multidrug-resistant, somewhere de novo sequencing technique is constantly mandatory because of the genetic multiplicity and vibrant genome reorganization take place [43,44]. WGM utilize single-molecule restriction examination to acquire results concerning the magnitude of the control splinter and their substantial positions beside the DNA strand [45]. Whole genome has been exploit in a number of functions, which comprised of phylogenetic analyses and genotyping of associated with microbial isolates [46,47], discovery of outsized genomic structural rearrangements or variations [48,49], and quality control or verification for genome sequences which were assembled [50,51]. Physical genome maps based a restriction also having the potential to be used as a model for the exact arrangement of NGS contigs and to assist concluding the spaces between mapped contig. Meanwhile the term "re-sequencing" used as sequencing tool, which refers to the act of sequencing several samples from the same group, where a suggested genome has been reported, and is used to support in the elucidation of the data composed and used for next generation sequencing advances. Such as, re-sequencing of human genomes has been used to discover of both mutations [52], and polymorphisms [53].

Targeted Region Specific Sequencing

Although most important development in NGS, sequencing data assembly, particularly from re-emerging pathogens or newly microorganisms, continue inhibited by the short of appropriate suggested sequences. De novo assemblage is the dominant technique to accomplish precise completed sequence, except multiple sequencing paired-end libraries or platforms are frequently essential to complete whole genome coverage. Onmus-Leone, Hang [54] introduced the technology to entire bacterial genome sequences assembled through assimilating shotgun Roche 454 pyro-sequencing through optical WGM. The WGRM (whole genome restriction map) is applied as the orientation to platform assembled sequence de novo contigs throughout a stepwise development (Figure 1). Outsized de novo contigs be positioned in the exact order and direction from side to side alignment to the WGRM. De novo contigs so as not could be aligned

to WGRM were combined into scaffolds with contig branching assembly evidence (Figure 1). These extensive scaffolds are subsequently aligned to the WGRM to discover the intersections to be removed along with the gaps and mismatches to be determined with unexploited contigs. The procedure was repeated in anticipation of a sequence with whole reportage and alignments through the full genome maps were accomplished. Exhausting this technique it is possible to achieve 100% WGRM coverage exclusive of a paired-end library [54].
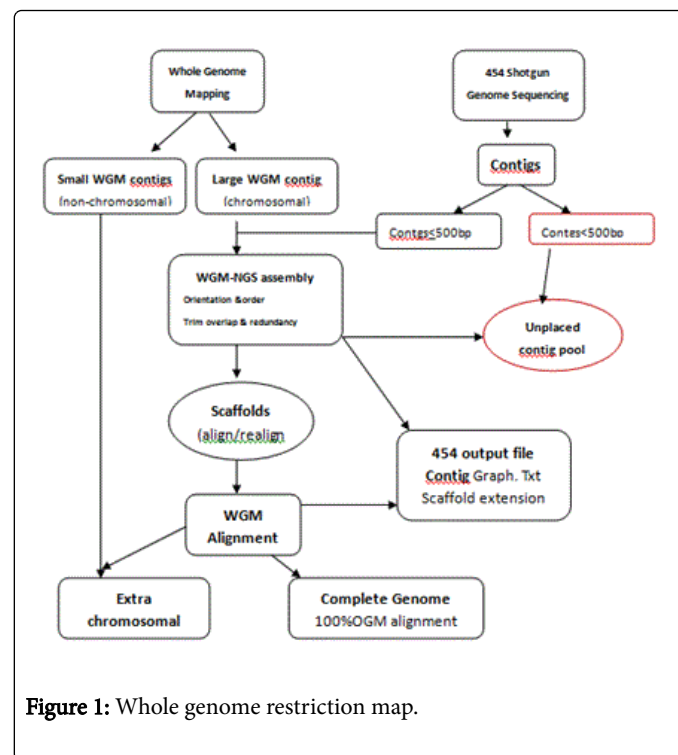

Figure 1: Whole genome restriction map.

Exome Sequencing

Exome is basically the protein coding content of the genetic code, which includes 1%–2% of the genome in all. While sequencers can read only so many bases per run, researchers sequencing exomes can produce more of them more quickly, at the greater resolution and the lower cost. Exome sequencing developed as DNA-enhancement tools and extremely corresponding nucleotide sequencing to discover altogether protein-coding modification in the genome. By regulating the extent of the experiments to the sequences of protein-coding, about five percent of human genome was sequenced. Collectively with developing unrestricted databanks of identified variants, exome sequencing permits for detection of genetic transformations in models were considered unsatisfactorily helpful for prior genetic studies [55]. Whole-genome and whole-exome sequencing were extremely successful in detecting the causes of Mendelian genetics in human. However, next-generation sequencing (NGS) has also progressed in identifying causes of genetic conditions [56,57]. The major challenge coupled along complete exome sequencing is the discovery of the disease-causing mutation(s) along with profuse hereditary candidate variants. Fuchs, Peeters-Scholte [58] described a number of approaches to handle this data wealth, counting association among control databases, growing number of patients and controls, and tumbling the genomic region under consideration throughout homozygosity

mapping. Recently, number of exceptional disarray of copper metabolism has been introduced with suspected, however unidentified monogenetic cause, as an attractive target for this approach. It is also anticipated that the application of these novel technologies will discover the basic deficiency in disorders illustrated, as well as in other genetic disorders of metal metabolism, in the next few years.

RNA Sequencing

Experimental miRNA investigation are frequently supplemented by bioinformatic techniques, are used to develop raw sequencing data, discover mature sequences, miRNA genes, targets and precursors to determine isoforms, and organize small RNAs into identified miRNA families [59,60]. These computational and experimental techniques not only permit for economical, qualitative and quantitative small RNAs analysis, they also produce more precise results in a limited time period [61]. Furthermore, in these sophisticated lab techniques, an opportunity of novel generation of bioinformatics methodologies has advanced emerged as necessary requirement to accommodate supplementary planned progress and enhancement of productivity. The HT-NGS is one of the enormous contests of genomic study.

Establishment of high-throughput tools and deep sequencing analysis has allowed the discovery of several miRNAs that are not preserved or are expressed in low levels, such as those found in Arabidopsis, wheat, rice, tomato and poplar [62]. RNA-sequencing (RNA-Seq) is extensively used for genomics study and exploring novel approaches to analyze the functional involvement of transcriptomes. Particularly Solexa/Illumina sequencing tools has many benefits as a revolutionary technology for transcriptome investigation, such as high exposure at a relatively low expenditure [63]. It has also been used to study transcriptomes in many plants, such as Arabidopsis, rice and berry [64,65]. Moreover, the Helicos BioSciences platform is suitable for applications that require quantitative insight in RNA-seq [66] or through RNA sequencing, as it sequences RNA templates directly without the required to translate them into cDNA's [67].

miRanalyzer

A large amount of algorithms has been used in order to develop outsized range of data [68,69]. Couple of years back, miRanalyzer was developed; a technique for the discovery of known microRNAs and prediction of novel ones in HTS experimentation with entirely redesigned and includes various novel features. The discovery of novel microRNAs is essentially significant approach as there are several species with very limited numbers were identified microRNAs. Hence, miRanalyzer has been developed as an online browsing tool, which realize all essential techniques for an inclusive analysis of deep-sequencing of small RNA molecules [70]. NGS platforms like Genome SequencerTM FLX or Genome Analyzer (Illumina Inc.) has become easily available for small RNA molecules sequencing, which allow both the new microRNA sequences, detection of expression levels at very rapid and sensitivity with economical for common researches. Though, every sequencing work yielded up to 3 Gbp of sequence data, and their data analysis represents a key bioinformatics solution and challenge [70]. Initially, the alignments are performed using the ultrafast short read aligner Bowtie [71,72] that contribute full color space support, allows incongruities in the alignment of the read to the genome and is more rapidly and memory capable than earlier applied alignment algorithm. Additional, the instrument covers 31 species and allocates simply addition of new ones. Third, the tool has no limitation on the number of loaded sequences for the prediction of new microRNAs,

and the instruction of the prediction models takes into account dissimilarity between animal and plant microRNAs [73]. Fourth, the implemented module, identifies the differential expression patterns of micro RNAs between two situations based on the DESeq package [74]. In addition, taking advantage of the information that several samples are required for this last module, the multiplication of the consensus sequences for predicted mature and precursor micro RNAs have also implemented. This will be helpful to evaluate the consistency of the extrapolations, i.e. micro RNAs predicted in diverse samples are probable to be efficient than those predicted in just single sample. In conclusion, a standalone form of the miRanalyzer tool that mechanized with modified local file-based database. Variation among animals and plants microRNAs have been used for the prophecy of models and disparity expression of identified and predicted microRNAs, among two situations could be intended. Furthermore, consent sequences of discovered mature and predecessor microRNAs can be acquired from various samples, which enhance the consistency of the expected microRNAs. At last, a stand-alone version of the miRanalyzer that is supported on a confined and simply modified database is also accessible for researches; this allows the client to have extra control on certain consideration as well as to use precise data, for example unpublished assemblies or other libraries that are inaccessible.

Whole Transcriptome

'Next-generation' sequencing (NGS) tools have been developed in a variety of systems for the investigation of the whole transcriptome for gene expression analysis [75,76]. NGS helps to examine the discovery of target genes and the deviations of their expression by discovering mRNA expression dissimilarity and functional annotation [77]. The comprehensive analysis of the transcriptome provides a significant platform to explore massive polymorphic molecular markers as a high quality resource of expressed sequence tag (EST) assortment [78].

RNA sequencing or Transcriptome sequencing is newly reported HTS technology that are able to generate millions of short cDNA reads in corresponding approach. RNA-seq can be used to verify sequences and a large quantity of transcripts, even at individual cellular level [79]. RNA sequencing has been extensively used in classification of transcriptome in model species, such as Arabidopsis and rice. Moreover, productively used in detection of long non-coding RNAs and alternatively spliced transcripts stresses responsive in Arabidopsis [80,81]. A general view of a transcriptome can be accessible by RNA sequencing, comprising new transcriptionally active regions and the specific position of transcription boundaries [82]. RNA-seq is particularly useful for investigation of transcriptomes of non-model species [83,84] as no previous understanding of transcript sequence is required.

Bioinformatic Challenges

Bioinformatics has become fundamental element of research and development in the biomedical sciences, and also plays an important function in deciphering genomic, proteomic and transcriptomic data produced by high throughput experimental tools, as systematizing information assembled from conventional biology [85]. Bioinformatics generally deals with four facets of analysis: protein structure prediction, DNA sequence analysis, proteomics, functional genomics and systems biology. High-throughput sequencing, with its rapidly declining expenditure and growing applications is substitute many other research technologies. Nonetheless, considerable confronts remain with NGS; these include data processing and storage. The very

large amount of data files Need massive quantity of data power (CPUs), Storage of Data and security/privacy (human samples) and the development of more proficient, strong and duplicated data analysis workflows, summarizing and Processing the huge quantity of data produced by HTS presents a nontrivial confront to bioinformatics [86]. One of the most triumphant output to homogenize HTS workflow was the expansion of the Sequence Alignment/Map (SAM) format for the storage of aligned sequencing reads, along with a analogous set of efficient programs working on SAM files [87].

Ethical Challenges

Sequencing of all genes is considered as a prognostic practice since analysis of such huge data can give unintentional findings and information about diseases in future [88]. Biomedical study is progressively having huge computational and intensive data, and "big data science" is transferred into the medical field. Unfortunately, regulators, ethicists, and policy-makers have hardly initiated to discover the social, legal and ethical issues raised by the multiplicity of analytical and computational advances in progress and under improvement in medicine and biology. The majority funding concerning huge data bioscience has paying attention on security, a significantly important deliberation, however not the single one. Along with the challenges raised by new computational techniques are investigation about security and safety consideration, justice, and how to attain appropriate informed consent. These expertises also raise a multitude of regulatory concerns that could pressure the prospects of translating new assays or computational technologies to the public health or clinical spheres [88]. In Norway, Biotechnology regulation and Predictive procedures are regulated and synchronized by the "bioteknologiloven". Although this situation raises several challenges, based on this law, written permission and genetic counseling before, during and after the predictive gene testing are demanded [89].

Bioinformatics Data Analysis Tools

High-throughput sequencing data was analyzed according to the three pipelines as shown in Figure 2. The objectives of these different methods are to determine the expression of miRNA with special techniques: the traditional method of reads alignment, and two latest techniques counting of isoforms which is autonomous of any database entries and seed investigation which focus on analysis to the bases that are mainly likely functionally significant [90]. In this regard several software's and procedures have been developed; including, 9 Analysis tools such as RMAP, BSMAP, mrsFAST, SOCS-B, BS-seeker, BRAT, MethylCoder, and Bismark as well as NGSmethPipe.
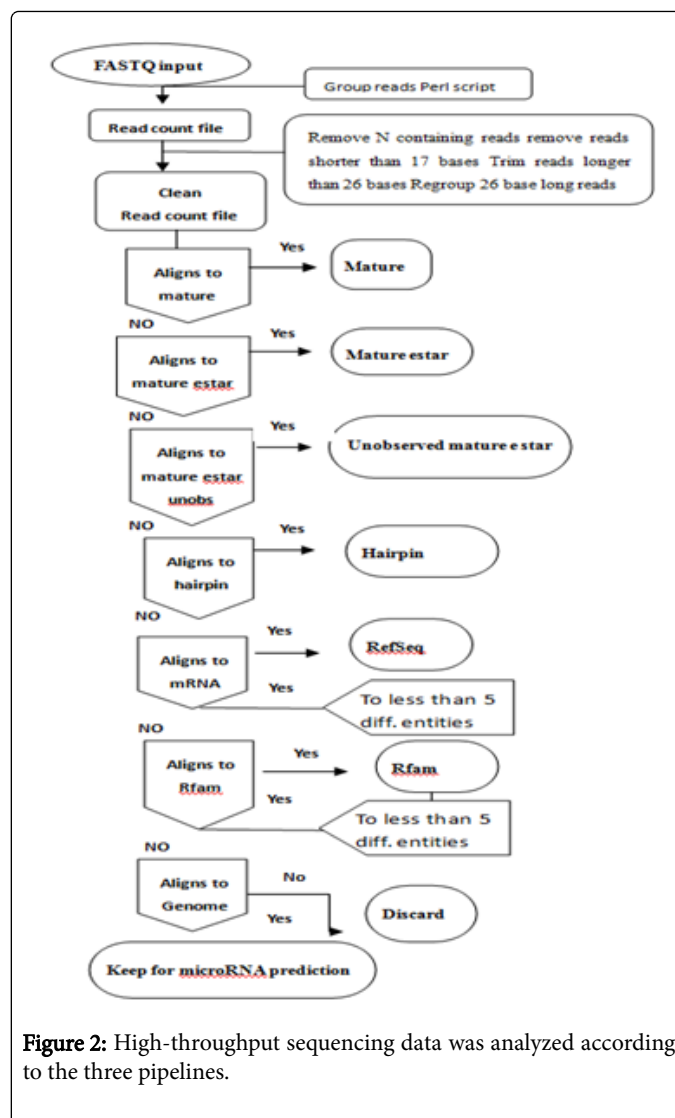

Figure 2: High-throughput sequencing data was analyzed according to the three pipelines.

Future Prospective

Third-generation sequencing technology

As technology moves forward, advancement made toward third generation sequencing tools are, being used which are comprised of real-time monitoring and Nanopore Sequencing of PCR activity throughout fluorescent resonant energy transfer. The benefit of these techniques consist of scalability, simplicity, with rising in DNA polymerase activities and products, with a reduction of miscalculation prone, and even more efficiently practicable with the ultimate goal of achieving precise real-time products [91].

The 3rd technology already being used in SMRT is to sequence number of DNA fragments in parallel on chip. The chip is comprised of aluminum deposited on top of glass microscope cover slip with a 100 nm-thick layer. There is an array of cylindrical wells 70–100 nm in diameter in the alluminium. Third generation sequencing technologies in present improvements over existing sequencing techniques are (i) Higher throughput; (ii) Longer read lengths to improve de novo assembly and allow to direct discovery of haplotypes and even

entire chromosome phasing; (iii) more rapid turnaround time; (iv) higher consensus accuracy to enable rare variant detection; (v) limited use of starting material (theoretically only a single molecule may be required for sequencing); and (vi) economical, wherever sequencing the human genome at high fold range for a lesser amount than $100 is now a realistic goal for the society [92].

Sequencing in diagnostics

Progress in DNA sequencing techniques has permitted complete examination of the genetics. Understandings from sequencing the transcriptomes, genomes or exomes, of diseased and healthy cells in longsuffering are already enabling enhanced analytical prognostication, classification, and therapy selection of several syndromes. Understanding the data acquired using new HTS-DNA tools, preference made in sequencing policies, and general issues in data analysis and genotype-phenotype correlation is vital if clinicians, geneticists, and pathologists understand the growing technical literature in this field, in all aspects [93]. The widespread genetic heterogeneity has a key apprehension for genetic counseling and molecular diagnosis. Whereas diverse approaches have been currently projected to optimize mutation detection, they either not succeed to perceive mutations in a many patients or done in a time consuming and expensive manner [94].

Competing Interests

The authors declared that they have no competing interests.

Acknowledgement

References

1. Chiu RW, Akolekar R, Zheng YW, Leung TY, Sun H, et al. (2011) Non-invasive prenatal assessment of trisomy 21 by multiplexed maternal plasma DNA sequencing: large scale validity study. BMJ 342: 7401.

2. Park SJ, SaitoAdachi M, Komiyama Y, Nakai K (2016) Advances, practice, and clinical perspectives in highthroughput sequencing. Oral Dis 22: 353-364.

3. Reuter JA, Spacek DV, Snyder MP (2015) High-throughput sequencing technologies. Mol Cell 58: 586-597.

4. Furey TS (2012) ChIP–seq and beyond: new and improved methodologies to detect and characterize protein–DNA interactions. Nat Rev Genet 13: 840-852.

5. Veltman JA, Brunner HG (2012) De novo mutations in human genetic disease. Nat Rev Genet 13: 565-575.

6. Treangen TJ, Salzberg SL (2011) Repetitive DNA and next-generation sequencing: computational challenges and solutions. Nat Rev Genet 13: 36-46.

7. Churko JM, Laird DW (2013) Gap junction remodeling in skin repair following wounding and disease. Physiology 28: 190-198.

8. Bentley DR, Balasubramanian S, Swerdlow HP, Smith GP, Milton J, et al. (2008) Accurate whole human genome sequencing using reversible terminator chemistry. Nature 456: 53-59.

9. Korlach J, Webb WW, Levene M, Turner S, Craighead HG, et al. (2008) Composition for nucleic acid sequencing. Google Patents.

10. Ansorge WJ (2009) Next-generation DNA sequencing techniques. New Biotechnol 25: 195-203.

11. Metzker ML (2009) Sequencing technologies—the next generation. Nat Rev Genet 11: 31-46.

12. Kircher M, Kelso J (2010) Highthroughput DNA sequencing–concepts and limitations. Bioessays 32: 524-536.

13. Wu B, Wang M, Ma Y, Yuan L, Lu S (2012) High-throughput sequencing and characterization of the small RNA transcriptome reveal features of novel and conserved microRNAs in Panax ginseng. PloS One 7: e44385.

14. Reinhart BJ, Weinstein EG, Rhoades MW, Bartel B, Bartel DP (2002) MicroRNAs in plants. Genes Dev 16: 1616-1626.

15. Kozomara A, Griffiths SJ (2011) miRBase: integrating microRNA annotation and deep-sequencing data. Nucleic Acids Res 39: 152-157.

16. Valiollahi E, Farsi M, Fevereiro P, Kakhki AM (2014) Bioinformatic characterization and expression analysis of miRNAs in Solanum lycopersicum. Plant Omics 7: 108-116.

17. Farazi TA, Brown M, Morozov P, Hoeve JJ, Ben-Dov IZ, et al. (2012) Bioinformatic analysis of barcoded cDNA libraries for small RNA profiling by next-generation sequencing. Methods 58: 171-187.

18. Sun G, Li M, Li G, Tian Y, Han R, et al. (2012) Identification and abundance of miRNA in chicken hypothalamus tissue determined by Solexa sequencing. Genet Mol Res 11: 4682-4694.

19. Crowhurst RN, Gleave AP, MacRae EA, Ampomah-Dwamena C, Atkinson RG, et al. (2008) Analysis of expressed sequence tags from Actinidia: applications of a cross species EST database for gene discovery in the areas of flavor, health, color and ripening. BMC Genomics 9: 351.

20. Newcomb RD, Crowhurst RN, Gleave AP, Rikkerink EH, Allan AC, et al. (2006) Analyses of expressed sequence tags from apple. Plant Physiol 141: 147-166.

21. Malone JH, Oliver B (2011) Microarrays, deep sequencing and the true measure of the transcriptome. BMC Biol 9: 34.

22. Sanger F, Air GM, Barrell B, Brown N, Smith M, et al. (1977) Nucleotide sequence of bacteriophage phi X174 DNA. Nature 265: 687-695.

23. Emrich CA, Tian H, Medintz IL, Mathies RA (2002) Microfabricated 384-lane capillary array electrophoresis bioanalyzer for ultrahigh-throughput genetic analysis. Anal Chem 74: 5076-5083.

24. Shendure JA, Porreca GJ, Church GM, Gardner AF, Hendrickson CL, et al. (2011) Overview of DNA sequencing strategies. Curr Protoc Mol Biol 7: 23.

25. Hert DG, Fredlake CP, Barron AE (2008) Advantages and limitations of nextgeneration sequencing technologies: A comparison of electrophoresis and nonelectrophoresis methods. Electrophoresis 29: 4618-4626.

26. Chang YJ, Chen CC, Ho JM, Chen CL (2012) De Novo Assembly of High-Throughput Sequencing Data with Cloud Computing and New Operations on String Graphs. Cloud Computing (CLOUD) Framework. BMC Genomics 13: S28.

27. Mulsant CMP, Gasser F (1998) Expressed sequence tags for genes: A review. Genet Sel Evol 30: 521-541.

28. Seki M, Shinozaki K (2009) Functional genomics using RIKEN Arabidopsis thaliana full-length cDNAs. J Plant Res 122: 355-366.

29. Garg R, Patel RK, Tyagi AK, Jain M (2011) De novo assembly of chickpea transcriptome using short reads for gene discovery and marker identification. DNA Res 18: 53-63.

30. Edery I, Chu LL, Sonenberg N, Pelletier J (1995) An efficient strategy to isolate full-length cDNAs based on an mRNA cap retention procedure (CAPture). Mol Cell Biol 15: 3363-3371.

31. Clepet C (2011) RNA captor: a tool for RNA characterization. PloS One 6: e18445.

32. Kim HA, Lim CJ, Kim S, Choe JK, Jo SH, et al. (2014) High-Throughput Sequencing and De Novo Assembly of Brassica oleracea var. Capitata L. for Transcriptome Analysis. PloS One 9: e92087.

33. Kumar S, Blaxter ML (2010) Comparing de novo assemblers for 454 transcriptome data. BMC Genomics 11: 571.

34. Martin J, Bruno VM, Fang Z, Meng X, Blow M, et al. (2010) Rnnotator: an automated de novo transcriptome assembly pipeline from stranded RNA-Seq reads. BMC Genomics 11: 663.

35. Hess M, Sczyrba A, Egan R, Kim TW, Chokhawala H, et al. (2011) Metagenomic discovery of biomass-degrading genes and genomes from cow rumen. Science 331: 463-467.

36. Ross EM, Moate PJ, Bath CR, Davidson SE, Sawbridge TI, et al. (2012) High throughput whole rumen metagenome profiling using untargeted massively parallel sequencing. BMC Genetics 13: 53.

37. Johansson H, Isaksson M, Sörqvist EF, Roos F, Stenberg J, et al. (2010) Targeted resequencing of candidate genes using selector probes. Nucleic Acids Res 39: e8.

38. Hu B, Xie G, Lo CC, Starkenburg SR, Chain PS (2011) Pathogen comparative genomics in the next-generation sequencing era: genome alignments, pangenomics and metagenomics. Brief Funct Genomics 10: 322-333.

39. Nagarajan N, Cook C, Di Bonaventura M, Ge H, Richards A, et al. (2010) Finishing genomes with limited resources: lessons from an ensemble of microbial genomes. BMC Genomics 11: 242.

40. Loman NJ, Misra RV, Dallman TJ, Constantinidou C, Gharbia SE, et al. (2012) Performance comparison of benchtop high-throughput sequencing platforms. Nature Biotechnol 30: 434-439.

41. Narzisi G, Mishra B (2011) Comparing de novo genome assembly: the long and short of it. PloS One 6: e19175.

42. Chain P, Grafham D, Fulton R, Fitzgerald M, Hostetler J, et al. (2009) Genome project standards in a new era of sequencing. Science 326: 236-237.

43. Hongoh Y, Toyoda A (2011) Whole-genome sequencing of unculturable bacterium using whole-genome amplification. High-Throughput Next Generation Sequencing 125: 25-33.

44. Köser CU, Holden MT, Ellington MJ, Cartwright EJ, Brown NM, et al. (2012) Rapid whole-genome sequencing for investigation of a neonatal MRSA outbreak. New Eng J Med 366: 2267-2275.

45. Neely RK, Deen J, Hofkens J (2011) Optical mapping of DNA: Singlemoleculebased methods for mapping genomes. Biopolymers 95: 298-311.

46. Ramirez MS, Adams MD, Bonomo RA, Centrón D, Tolmasky ME (2011) Genomic analysis of Acinetobacter baumannii A118 by comparison of optical maps: identification of structures related to its susceptibility phenotype. Antimicrob Agents Chemother 55: 1520-1526.

47. Clarridge JE, Harrington AT, Roberts MC, Soge OO, Maquelin K (2012) Impact of strain typing method on assessing the relationship of paired nares and wound isolates of methicillin-resistant Staphylococcus aureus. J Clin Microbiol 51: 224-231.

48. Saunders M, Wu G, Abuoun M, Pan Z, Anjum M, et al. (2011) Optical genetic mapping defines regions of chromosomal variation in serovars of S. enterica subsp. enterica of concern for human and animal health. Epidemiol Infect 139: 1065-1074.

49. Turner PC, Yomano LP, Jarboe LR, York SW, Baggett CL, et al. (2012) Optical mapping and sequencing of the Escherichia coli KO11 genome reveal extensive chromosomal rearrangements, and multiple tandem copies of the Zymomonas mobilis pdc and adhB genes. J Indust Microbiol Biotechnol 39: 629-639.

50. Giongo A, Tyler HL, Zipperer UN, Triplett EW (2010) Two genome sequences of the same bacterial strain, Gluconacetobacter diazotrophicus PAl 5, suggest a new standard in genome sequence submission. Standards Genomic Sci 2: 309.

51. Yoo M, Kim D, Choi KY, Chae JC, Zylstra GJ, et al. (2012) Draft genome sequence and comparative analysis of the superb aromatic-hydrocarbon degrader Rhodococcus sp. strain DK17. J Bacteriol 194: 4440.

52. Mardis ER, Ding L, Dooling DJ, Larson DE, McLellan MD, et al. (2009) Recurring mutations found by sequencing an acute myeloid leukemia genome. New Eng J Med 361: 1058-1066.

53. Singh DP, Prabha R, Rai A, Arora DK (2012) Bioinformatics-assisted microbiological research: tasks, developments and upcoming challenges. Am J Bioinformat 1: 10.

54. Onmus-Leone F, Hang J, Clifford RJ, Yang Y, Riley MC, et al. (2013) Enhanced de novo assembly of high throughput pyrosequencing data using whole genome mapping. PloS One 8: e61762.

55. Kota DJ, Seed J, Johnson J (2013) Exome Sequencing – Research and Clinical Applications.

56. www.edgebio.com.

57. Vissers LE, de Ligt J, Gilissen C, Janssen I, Steehouwer M, et al. (2010) A de novo paradigm for mental retardation. Nat Genet 42: 1109-1112.

58. Need AC, Shashi V, Hitomi Y, Schoch K, Shianna KV, et al. (2012) Clinical application of exome sequencing in undiagnosed genetic conditions. J Med Genet 49: 353-361.

59. Fuchs SA, Peeters-Scholte CM, de Barse MM, Roeleveld MW, Klomp LW, et al. (2012) Increased concentrations of both NMDA receptor co-agonists D-serine and glycine in global ischemia: a potential novel treatment target for perinatal asphyxia. Amino Acids 43: 355-363.

60. Motameny S, Wolters S, Nürnberg P, Schumacher B (2010) Next generation sequencing of miRNAs–strategies, resources and methods. Genes 1: 70-84.

61. Dai X, Zhuang Z, Zhao PX (2011) Computational analysis of miRNA targets in plants: current status and challenges. Brief Bioinformat 12: 115-121.

62. Lhakhang TW, Chaudhry MA (2012) Interactome of radiation-induced microRNA-predicted target genes. Compar Funct Genom 2012: 264-270.

63. Wei B, Cai T, Zhang R, Li A, Huo N, et al. (2009) Novel microRNAs uncovered by deep sequencing of small RNA transcriptomes in bread wheat (Triticum aestivum L.) and Brachypodium distachyon (L.) Beauv. Funct Integr Genomics 9: 499-511.

64. Costa V, Angelini C, Ciccodicola A (2010) Uncovering the complexity of transcriptomes with RNA-Seq. J Biomed Biotechnol 2010: 210-229.

65. Lister R, O'Malley RC, Tonti-Filippini J, Gregory BD, Berry CC, et al. (2008) Highly Integrated Single-Base Resolution Maps of the Epigenome in Arabidopsis. Cell 133: 523-536.

66. Zenoni S, Ferrarini A, Giacomelli E, Xumerle L, Fasoli M, et al. (2010) Characterization of transcriptional complexity during berry development in Vitis vinifera using RNA-Seq. Plant Physiol 152: 1787-1795.

67. Lipson D, Raz T, Kieu A, Jones DR, Giladi E, et al. (2009) Quantification of the yeast transcriptome by single-molecule sequencing. Nat Biotechnol 27: 652-658.

68. Ozsolak F, Platt AR, Jones DR, Reifenberger JG, Sass LE, et al. (2009) Direct RNA sequencing. Nature 461: 814-818.

69. Pantano L, Estivill X, SeqBuster ME (2010) A bioinformatic tool for the processing and analysis of small RNAs datasets, reveals ubiquitous miRNA modifications in human embryonic cells. Nucleic Acids Res 38: e34-e40.

70. Ronen R, Gan I, Modai S, Sukacheov A, Dror G, et al. (2010) miRNAkey: a software for microRNA deep sequencing analysis. Bioinformat 26: 2615-2616.

71. Hackenberg M, Sturm M, Langenberger D, Falcon-Perez JM, Aransay AM (2009) miRanalyzer: a microRNA detection and analysis tool for next-generation sequencing experiments. Nucleic Acids Res 37: 68-76.

72. Langmead B, Trapnell C, Pop M, Salzberg SL (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol 10: R25.

73. Hackenberg M, Rodríguez-Ezpeleta N, Aransay AM (2011) miRanalyzer: an update on the detection and analysis of microRNAs in high-throughput sequencing experiments. Nucleic Acids Res 39: 132-138.

74. Millar AA, Waterhouse PM (2005) Plant and animal microRNAs: similarities and differences. Funct Integr Genomics 5: 129-135.

75. Anders S, Huber W (2010) Differential expression analysis for sequence count data. Genome Biol 11: R106.

76. Cohen D, Bogeat-Triboulot MB, Tisserant E, Balzergue S, Martin-Magniette ML, et al. (2010) Comparative transcriptomics of drought responses in Populus: a meta-analysis of genome-wide expression profiling in mature leaves and root apices across two genotypes. BMC Genomics 11: 630.

77. Roberts GC, Smith CW (2002) Alternative splicing: combinatorial output from the genome. Curr Opin Chem Biol 6: 375-383.

78. Lee HR, Bae IH, Park SW, Kim HJ, Min WK, et al. (2009) Construction of an integrated pepper map using RFLP, SSR, CAPS, AFLP, WRKY, rRAMP, and BAC end sequences. Mol Cells 27: 21-37.

79. Ahn YK, Tripathi S, Kim JH, Cho YI, Lee HE, et al. (2014) High-throughput microsatellite marker information from Capsicum annuum varieties Mandarin and Blackcluster next-generation sequence data. Scientia Horticulturae 170: 123-130.

80. Tang F, Barbacioru C, Wang Y, Nordman E, Lee C, et al. (2009) mRNA-Seq whole-transcriptome analysis of a single cell. Nat Meth 6: 377-382.

81. Filichkin SA, Priest HD, Givan SA, Shen R, Bryant DW, et al. (2010) Genome-wide mapping of alternative splicing in Arabidopsis thaliana. Genome Res 20: 45-58.

82. Zhu QH, Stephen S, Taylor J, Helliwell CA, Wang MB (2014) Long noncoding RNAs responsive to Fusarium oxysporum infection in Arabidopsis thaliana. New Phytologist 201: 574-584.

83. Wilhelm BT, Marguerat S, Goodhead I, Bähler J (2010) Defining transcribed regions using RNA-seq. Nature protocols 5: 255-266.

84. Feng C, Chen M, Xu CJ, Bai L, Yin XR, et al. (2012) Transcriptomic analysis of Chinese bayberry (Myrica rubra) fruit development and ripening using RNA-Seq. BMC Genomics 13: 19.

85. Xia Z, Xu H, Zhai J, Li D, Luo H, et al. (2011) RNA-Seq analysis and de novo transcriptome assembly of Hevea brasiliensis. Plant Mol Biol 77: 299-308.

86. Al-Haggar M, Khair-Allaha B, Islam M, Mohamed A (2013) Bioinformatics in High Throughput Sequencing: Application in Evolving Genetic Diseases. J Data Mining Genom Proteom 4: 21.

87. Zhang Z, Leipzig J, Sasson A, Angela MY, Perin JC, et al. (2013) Efficient digest of high-throughput sequencing data in a reproducible report. BMC Bioinformat 14: S3.

88. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, et al. (2009) The sequence alignment/map format and SAMtools. Bioinformatics 25: 2078-2079.

89. Holdhus R (2011) High Throughput Sequencing applications. Bioinfo.

90. Ossorio PN (2014) The Ethics of Translating HighThroughput Science into Clinical Practice. Hastings Center Report 44: 8-9.

91. Stokowy T, Eszlinger M, Fujarewicz K, Jarz B, Paschke R, et al. (2014) Analysis options for high-throughput sequencing in miRNA expression profiling. BMC Res Notes 7: 144.

92. Diaz-Sanchez S, Hanning I, Pendleton S, D'Souza D (2013) Next-generation sequencing: The future of molecular genetics in poultry production and food safety. Poultry Sci 92: 562-572.

93. Schadt EE, Turner S, Kasarskis A (2010) A window into third-generation sequencing. Human Mol Genet 19: 227-240.

94. Boyd SD (2013) Diagnostic applications of high-throughput DNA sequencing. Annual Review of Pathology: Mech Dis 8: 381-410.

95. Redin C, Le Gras S, Mhamdi O, Geoffroy V, Stoetzel C, et al. (2012) Targeted high-throughput sequencing for diagnosis of genetically heterogeneous diseases: efficient mutation detection in Bardet-Biedl and Alström syndromes. J Med Genet 49: 502-512.