

## **Next-generation sequencing technologies as emergent tools and their challenges in viral diagnostic.**

**Desh Deepak Singh\***

Amity institute of Biotechnology, Amity University, Rajasthan, India

### **Abstract**

**Next-Generation Sequencing (NGS) has the range of high-throughput quickly adapted into the countless facet of viral diagnostic research. This method has been extensively applying in full genome sequencing, genetic diversity identification, transcriptomic routine diagnostic work patient care and management, and the new understanding of the interactions between viral and host transcriptome, to promote virus research. An exciting era of viral exploration has begun and will set us new challenges to understand the role of newly discovered viral diversity in both disease and health. In this review, discuss about the preparation of viral nucleic acid templates for sequencing, assay formats underlying next-generation sequencing systems, methods for imaging and base calling, quality control, data processing pipelines and bioinformatics approaches for sequence alignment, variant calling and suggestions for selecting suitable tools. Also discuss of the most important advances that the new sequencing technologies have brought to the fields of clinical virology and challenges behind it.**

**Keywords:** Next generation sequencing, Virology, Bioinformatics, Transcriptome, Viral diversity, Sequencing depth.

*Accepted on February 17, 2018*

### **Introduction**

Many viruses show the high degree of genetic diversity continuously, due to a large number of evolutionary changes, short generation times, globalization, climate changes, large population sizes, and increased number of immunocompromised people [1]. The human is getting infected with a cumulative problem of viral diseases caused by the appearance of novel unidentified viruses and emergence of new infections [2]. Medically important immune immunodeficiency virus and hepatitis virus, Influenza like illness- causing viruses and severe acute respiratory illness causing viruses and high genetic heterogeneity causing virus in the patient [3]. Assessing of intra-host viral genetic diversity is really very important to know the evolutionary sequences of viruses, for manipulative real vaccines, and for the achievement of antiviral therapy [4,5].

Sometimes viral etiology cannot be characterized by traditional culture and molecular methods [6]. Scarce techniques, such as depictive modification investigation or arbitrary sequencing of plasmid libraries of nuclease resilient drivels of viral genomes [7], take lead in the previous to the discovery of several viruses, counting polyomavirus [8]. Human herpes virus type [9], Human GB [10] human parvovirus [11].

Need to advance approaches for the genetic characterization of suspected viral pathogens or new viruses [12]. So, NGS techniques signify an influential tool which can be useful to metagenomics founded approaches for the identification of novel pathogenic viruses in patients care [13]. In this review, I

will discuss an overview of NGS challenge, opportunity in virology, specifically in virus discovery, genome sequencing and transcriptomics.

### **Next-generation Sequencing**

NGS is capable of sequencing large numbers of different DNA sequences in a single reaction [14]. All Next-Generation Sequencing (NGS) technologies monitor the sequential addition of nucleotides to immobilized and spatially arrayed DNA templates, but differ substantially in how these templates are generated and how they are interrogated to reveal their sequences [15]. Current NGS platforms may show significant differences and their different characteristics are mentioned in (Table 1).

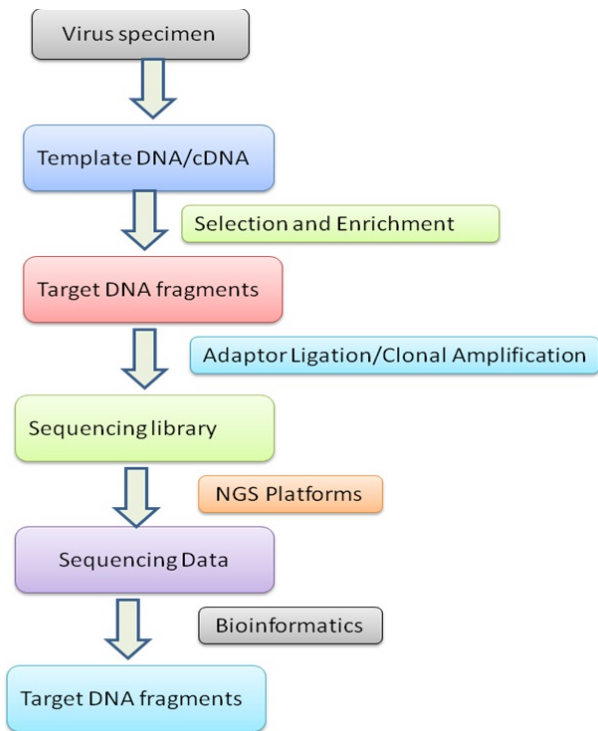
#### **Choice of platform**

Important factors are included to choose of the machine, such as the size of the genome, G+C content, as well as the depth of coverage and accuracy required [16]. It is, therefore, most important to take advice from local service providers according to your requirement.

### **Methods**

The data of NGS for viral diversity analysis depends significantly on the quality and preparation of the samples [17]. The high-quality of protocols castoff for genome extraction and removal of unwanted RNA and DNA from other sources

like host cells depends on experiment handling [18]. To diminish experimental variations, individual step needs cautious attention from preparation to sequence data analysis (Figure 1).



**Figure 1.** Workflow of next-generation sequencing.

### Sample preparation

Nucleic acids may be extracted from whole blood, plasma, serum, throat swabs, cerebral spinal fluid, virus-infected supernatants, and other cell-free body fluids by using a nucleic acid extraction kit or manually by triazole methods using omitting RNase or DNase digestion according to the manufacturer's recommended protocols [19,20]. RNA is bound to an advanced silica gel membrane under optimal buffering conditions [21]. A simple two-step washing protocol ensures that PCR inhibitors such as proteins or divalent cations are completely removed, leaving high-quality RNA to be eluted in Milli-Q water [22]. Nucleic acid extraction from viral samples brings with it the probable extra contest of low or variable viral titers [23]. Users are advised to low viral titers must find a method that is easy to use and that ensures high quantity for NGS [24].

### Library preparation

The clonal amplification of each DNA fragment is performed by bridge amplification or emulsion PCR is necessary to generate sufficient copies of sequencing template [25]. The fragment libraries are obtained by annealing platform-specific adaptors to fragments generated from a DNA source of interest, such as genomic DNA, double-stranded cDNA, and PCR amplicons [26]. The presence of adapter sequences enables selective clonal amplification of the library molecules

[27]. Therefore, no bacterial cloning step is required to amplify the template fragment in a bacterial intermediate, as is performed in traditional sequencing approaches; the adapter sequence also contains a docking site for the platform-specific sequencing primers [28]. The conventional method for DNA library preparation contains of 4 steps (I) Fragmentation of DNA (II) End overhaul of fragmented DNA (III) Ligation of adapter sequences (not for single-molecule sequencing applications) (IV) Optional library amplification [29]. At present different method are used to produce fragmented genomic DNA, such as sonication, nebulization, enzymatic digestion, and hydrodynamic shearing. Each method has specific compensation and limitations [30]. Enzymatic digestion is simple and rapid, but it is frequently tricky to precisely control the fragment length distribution [31]. In addition, this method is likely to carry in biases regarding the representation of genomic DNA [32]. The further three techniques make use of physical methods to start dual strand breaks into DNA, which are supposed to arise randomly resulting in an unbiased representation of the DNA in the library [33]. The subsequent DNA fragments size qualitative analysis can be done by agarose gel electrophoresis [34].

Following fragmentation, the DNA sections must be repaired to generate blunt-ended, 5'-phosphorylated DNA ends compatible with the sequencing platform-specific adapter ligation strategy [35]. The library generation efficiency is directly dependent on the efficiency and accuracy of these DNA end-repair steps [36].

The end-repair mix converts 5'- and 3'-overhanging ends to 5'-phosphorylated blunt-ended DNA [37]. In most cases, the end repair is proficient by manipulating the 5'-3' polymerase and the 3'-5' exonuclease activities of T4 DNA polymerase, although T4 Polynucleotide Kinase confirms the 5'-phosphorylation at blunt-end DNA fragments, preparing these fragments for subsequent adapter ligation [38]. The blunt-end DNA fragments can either directly be cast-off for adapter-ligation, or essential the adding of a single A overhang at the 3' ends of the DNA fragments to simplify succeeding ligation of platform-specific connectors through well-suited solitary T overhangs [39]. Classically, this A-addition stage is catalysed by Klenow Fragment (minus 3' to 5' exonuclease) or other polymerases with terminal transferase activity [40]. Ligase repaired library fragments, followed by reaction cleanup and DNA size selection to remove free library adapters [41]. The approaches for size selection of library comprise agrose gel separation, the usage of magnetic beads, or progressive column-based refinement methods [42]. Adapter-dimers that can occur during the ligation and will subsequently be co-amplified with the adapter-ligated library fragments must be depleted from the libraries prior to sequencing, as they reduce the capacity of the sequencing platform for real library fragments and reduce sequencing quality [43]. Some sequencing platforms require a narrow distribution of library fragments for optimal results, which in many cases can only be achieved by excising the respective fragment section from the gel. This can also serve to deplete adapter-dimmers [44].

Fragment DNA libraries should be competent and quantified [45]. During library amplification stage, high-quality DNA polymerases are used to either produce the complete adapter sequence desirable for consequent clonal amplification and binding of sequencing primers with overlapping PCR primers, and/or to produce sophisticated vintages of the DNA libraries [46]. Optimal library amplification requires DNA polymerase with high fidelity and minimal sequence bias [47].

To enable efficient use of the sequencing capacity, sequencing libraries generated from different samples can be pooled and sequenced in the same sequencing run [48]. This is enabled by ligation DNA fragments to adaptors with characteristic barcodes, i.e., short stretches of nucleotide sequences that are distinct for each sample [49].

A complete sequencing library can subsequently be constructed by limited rounds of the PCR amplification of such tagged DNA fragments, limiting handling steps and saving time [50]. However, libraries generated using *in vitro* transposition may show higher sequence bias compared to those generated using conventional methods [51].

### Library quality control for NGS

A high-quality library is a key to successful NGS [52]. Library construction includes complex steps, such as fragmenting the

sample, repairing ends, adenylation of ends, ligation of adapters, and amplifying the library [53]. Monitoring of each step is needed to perform very carefully, including examination of the size of library after sample fragmentation and a size and concentration checked after ligation of adapters [54]. Library authentication helps by way of the final library excellence control step, which examines the library size and quantity [55].

### Sequencing

To accomplish nucleic acid sequence from the amplified libraries, the library fragments act as a template [56]. The sequencing occurs from beginning to end a cycle of washing and flooding the fragments with the recognized nucleotides in a sequential order [57]. As nucleotides incorporate into the growing DNA strand, they are digitally recorded as sequence [58]. The PGM and the MiSeq each rely on a slightly different mechanism for detecting nucleotide sequence information [59]. The PGM performs semiconductor sequencing that relies on the detection of pH changes induced by the release of a hydrogen ion upon the incorporation of a nucleotide into a growing DNA strand [60]. Mi-sequence, is based on the detection of fluorescence produced by the incorporation of fluorescently labeled nucleotides into the growing strand of DNA [61,62].

**Table 1.** Review of current next-generation technologies.

Machine	Method	Sequencing chemistry	Read length (bp)	Sequencing speed/h	Maximum output per run	Accuracy (%)	Error rate*	Main source of error
454 FLX	Emulsion PCR	Pyrosequencing	400-700	13 Mbp	700 Mbp	99.9	$10^{-2}$ - $10^{-4}$	Intensity cutoff, homopolymers, amplification, mixed beads
Illumina (Illumina)	Bridge PCR	Reversible terminators	100-300	25 Mbp	600 Gbp	99.9	$10^{-2}$ - $10^{-3}$	Homopolymers, phasing, nucleotide labeling, amplification, low coverage of AT rich regions
Solid (Life technologies)	Emulsion PCR	Ligation	75-85	21-28 Mbp	80-360 Gbp	99.9	$10^{-2}$ - $10^{-3}$	Phasing, nucleotide labeling, signal degradation, mixed beads, low coverage of AT rich regions
Helicos (Helicos Biosciences)	No amplification Single molecule	Reversible terminators	25-55	83 Mbp	35 Gbp	97	$10^{-2}$	Polymerase employed, molecule loss, low intensities
Ion torrent PGM	Emulsion PCR	Detection released H	of 100-400	25 Mb-16 Gbp	100 Mb-64 Gbp	99	$3 \times 10^{-2}$	Homopolymers, amplification

**Table 2.** Tools for next-generation sequencing data analysis.

Category	Tool	Reference
miRNA prediction	miRDeep/miRDeep2	[63]
Alignment/mapping	FastQC	[64]
De novo assembly	Velvet	[65]
Quality Viewer	qrc	[66]

Quality control	FASTQ Quality Filter	[67]
Adapter trimming	cutadapt	[68]
Alignment/mapping	Bowtie/Bowtie2	[69]
Counting reads per transcript	HTSeq	[70]
Normalization, bias correction, and statistically testing differential expression	DESeq	[71]

Alignment/mapping	MAQ	[72]
SNV detection	VarScan/VarScan2	[86]
Structural variation detection	PEMer	[87]

online and commercial software is cast-off for the purpose of NGS data analysis as shown in Table 2 [81].

### Read alignment

The orientation-created assemblage is employed to bring into line few hundred or thousand millions of alignment to an existing reference genome [82]. MAQ contrivance remains founded on the idea of a spread-out kernel indexing approach to map recites with orientation arrangement [83]. BFAST tool container usage for its speed and accuracy on mapping [84]. Novo align uses the Needle man-Wunsch procedure and affine gap consequences to the invention the universally optimal alignment. Burrows-Wheeler Aligner (BWA) short algorithm that queries short reads up to ~200 bp with a high sensitivity rate and queries long reads with a high error rate [85]. SOAP3, the latest version of SOAP supports Graphics Processing Unit (GPU)-based corresponding orientation and within 30 s for a one-million-read configuration onto the human orientation genome [86].

### De novo assembly

The *de novo* methods mainly for grouping short reads into significant contigs and accumulating these contigs addicted to supports to rebuild the unique genomic DNA for novel species [87]. The crucial challenge of *de novo* assembly is that the read length is shorter than recurrences in the genome [88]. Velvet is practical by the postponement of valuable graph popularization to reduce the path difficulty [89].

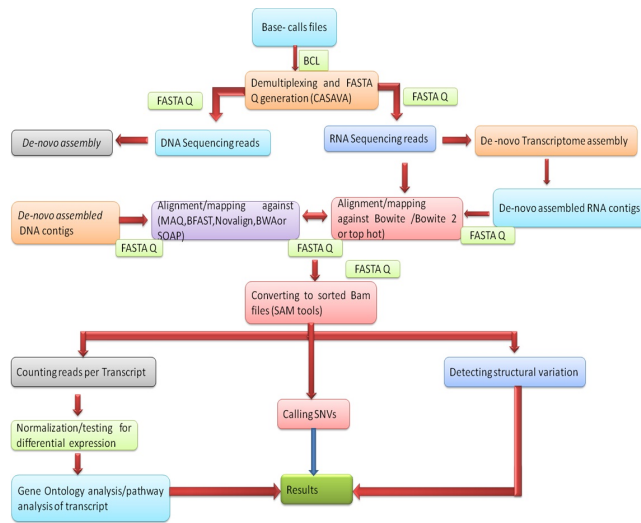
After gathering the sequence, the next step in systematic cylinders is using a tool to call SNVs for documentation of hereditary alternatives. GATK breaks re-alignment supplements/removals (indels), achieves dishonorable excellence recalibration, noises genotypes, and differentiates factual separating difference by machine learning to regulate and genotype differences among manifold samples. SAM and VarScan/VarScan2tools computes genotype likelihood to call SNVs or short indels [90], employs experiential approaches and a numerical test to detect SNVs and indels. Somatic Sniper [91] and Joint SNVMix [92], use the genotype prospect typical of MAQ and two probabilistic graphical models, individually, to measure the possibility of the variances among genotypes [93].

### NGS applications to virology

NGS applications to clinical virology briefly described here.

**Virus discovery (metagenomics):** Detection of novel viruses from clinically important samples in human and animal diseases, e.g. a new arenavirus tangled in resettle-related disease collections, [94] and the new Ebola virus Bundi-ubugyo [95], and the characterization of a viral etiology of an outbreak of a disease [96]. NGS is also useful in the environment [97], in animals [98,99], and in humans [100].

**Whole viral genome reconstruction:** Whole genome sequencing can be done by shotgun metagenomic sequencing



**Figure 2.** Steps for analytic strategies of DNA-seq and RNA-seq.

### Imaging

Throughout each run after NGS, more than a huge image data are generated and transformed to FASTQ format files [70]. Image examination usages raw images to trace clusters, export the positions and intensity and evaluate the noise for each cluster [71]. The base-calling stage recognizes the preparation of base reads from each cluster and filters undefined or low-quality reads [72]. In case of multiple samples identification, each sample analysed by its individual index sequences called bar-codes [73].

### Bioinformatics

Once the data is mapped onto the reference genome by different processors, and then aggregated by a head computing node to provide the final mapped genomic sequences which are desired as shown in Figure 2 [74]. Mapping of the data depending on the size of the genome that needs to be mapped [75]. For reference mapping, we can use a pre-existing reference genome sequence [76]. Information produced from a bench-top sequencer can be effortlessly examined by means of 96-256 GB RAM to achieve the mapping of sequence data, for full-length genome size is steadily advanced about 3 GB versus the 40-60 MB and a few MB are obligatory in the case of a steady clinical gene panel [77]. Denovo sequencing is mandatory lengthier read lengths and larger analysis to perform the overlapping process resourcefully [78]. Denovo read mapping is a considerable additional computationally concentrated course associated to the reference mapping procedure [79]. Here is a huge network of bioinformatics software obtainable for the resolution of examination of DNA sequencing data [80]. At present-day, a collection of free

of random libraries and shotgun sequencing of full genome amplicons, this type of method has been applied for sequencing of pandemic influenza virus [101] HIV [2], human herpes viruses [12], and other viruses, in most of the investigations of genetic diversity of intra-host virus variability has been observed.

**Characterization of intra-host variability:** NGS has been widely used for the characterization of intra-host variability of influenza virus [101], HCV [102], and HIV [2]. Due to high replication ability of the replication [12], though a virus commonly administrates in the early stages of disease, throughout the following stages a large number of mutations emerges naturally [95]. These changes are rapidly improved [98]. The highly variable district of classifications inside an expected host, mentioned to as quasi-species, allows a viral population to rapidly adapt to dynamic environments and evolve resistance to vaccines and antiviral drugs [101]. Most of these studies have been based on Ultra-Deep Sequencing (UDS) of PCR-generated amplicons spanning the genomic regions of interest [100].

### ***Epidemiology of viral infections and viral evolution***

NGS is existence used to examine the epidemiology of viral infections and viral progress, addressing issues such as viral super infection, which occurs when a previously infected individual acquires a new distinct strain [103], tracing the evolution and spread of viral strains, such as the emergence, evolution and worldwide spread of viruses, tracing the spread of viruses among persons [66], or showing the progress of viruses within the host and the machine of immune outflow, composed with replication suitability, such as in the case of HIV [2], and HCV infection [102].

### ***Quality control of live-attenuated viral vaccine***

Hereditary instability of RNA viruses may lead to the accumulation of infectious regress ants throughout the production of living viral vaccines, necessitating detailed excellence controller to confirm vaccine protection [5]. NGS approaches have been expected as challenges for exhaustive maintenance inherent reliability of live viral vaccines.

### ***Challenges***

Various problems can require exploit since the scientific public in contradiction of practical contests to differentiate between de-novo transcript and post-transcriptional modifications [10].

The huge datasets formed by the diverse NGS technologies derived from their own challenges [13]. In addition, storing planetary, it will be significant to have correct arrangement purposes to be able to map cDNA reads onto the genome and to eliminate poor-quality nucleotide bases [9]. Here might else be problems with repetitive sequences and homopolymeric regions at the 5' or 3' ends of cDNA reads, which can confuse 5' RACE and 3' RACE experiments [10]. This is a common problem with 454 FLX sequencing as this is identified to absence precision at homopolymer regions [62]. Huge datasets

will allow more accurate identification of transcript levels and associated strategies, but remain the risk of data overflow [78]. Finally, conception, examination and clarification will need substantial points of knowhow, and might similarly necessitate program design services [82]. Conception might be realized with the afore stated ARTEMIS [72] and combined genome Affymetrix, but then again LASERGENE (DNASTAR) also offer modules optimized for RNA-seq analyses [81].

### ***Conclusion***

Emerging tools for NGS analysis of viral populations are required. As new sequencing technologies continue to emerge, other future bioinformatic challenges will include developing algorithms for aligning very long reads, and coping with the unique error profiles of each sequencing technology. Investigative virology is one of the greatest positive requests for NGS and thrilling consequences have been attained in the finding and classification of new viruses, detection of unexpected viral pathogens in viral sample, ultrasensitive intensive care of antiviral drug discovery, examination of viral multiplicity, development and transmission, and human virome. With the decrease of costs and development of turnaround period, these methods will possibly develop important analytical tools trendy clinical procedures samples. The possible application of these methods in documentation and confinement agendas can efficiently recover the competence and dependability of investigation authentication and in monitoring viral diseases at world level.

### ***References***

1. Scheuch M, Hoper D, Beer M. RIEMS a software pipeline for sensitive and comprehensive taxonomic classification of reads from metagenomics datasets. BMC Bioinform 2015; 45: 69.
2. Gianella S, Delport W, Pacold ME. Detection of minority resistance during early HIV- 1 infection: natural variation and Spurious detection rather than transmission and evolution of multiple viral variants. J Viro 2011; 85: 8359-8367.
3. Casbon JA, Osborne RJ, Brenner S, Lichtenstein CP. A method for counting PCR template molecules with application to next- generation sequencing. Nucl Acids Res 2011; 39: 81.
4. Pyrc K, Jebbink MF, Berkhout B. Detection of new viruses by VIDISCA. Virus discovery based on cDNA-amplified fragment length polymorphism. Meth Molecul Bio 2008; 454: 73-89.
5. Radford AD, Chapman D, Dixon L. Application of next- gene rations sequencing technologies in virology. J Gen Virol 2012; 93: 1853-1868.
6. Bexfield N, Kellam P. Metagenomics and the molecular identification of novel viruses. Vet J 2011; 190: 191-198.
7. Dean FB, Nelson JR, Giesler TL, Lasken RS. Rapid amplification of plasmid and phage DNA using phi29 DNA polymerase and multiply-primed rolling circle amplification. Gen Res 2001; 11: 1095-1099.

8. Gaynor AM, Nissen MD, Whiley DM. Identification of a novel polyomavirus from patients with acute respiratory tract infections. *PLoS Pathogen* 2010; 3.
9. Szpara ML, Parsons L, Enquist LW. Sequence variability in clinical and laboratory isolates of herpes simplex virus 1 reveals new mutations. *J Virol* 2010; 84: 5303- 5313.
10. Simons JN, Pilot-Matias TJ, Leary TP. Identification of two flavivirus-like genomes in the GB hepatitis agent. Chalmers. *Proc Natl Acad Sci USA* 1995; 92: 3401-3405.
11. Allander T, Tammi MT, Eriksson M. Cloning of a human parvovirus by molecular screening of respiratory tract samples. *Proc Natl Acad Sci USA* 2005; 102: 12891-12896.
12. Cheval J, Sauvage V, Frangeul L. Evaluation of high throughput sequencing for identifying known and unknown viruses in biological samples. *J Clin Micro* 2011; 49: 3268-3275.
13. Yozwiak NL, Skewes-Cox P, Stenglein MD. Virus identification in unknown tropical febrile illness cases using deep sequencing. *PLoS Negl Trop Dis* 2012; 6: 1485.
14. Metzker ML. Sequencing technologies the next generation. *Nat Rev Genet* 2010; 11: 31-46.
15. Wang D, Urisman A, Liu YT. Viral discovery and sequence recovery using DNA microarrays. *PLoS Biol* 2003; 1: 257-260.
16. Shendure J, Ji H. Next-generation DNA sequencing. *Nat Biotechnol* 2008; 26: 1135-1145.
17. Barzon L, Lavezzo E, Militello V. Applications of next-generation sequencing technologies to diagnostic virology. *Inter J Mole Sci* 2011; 12: 7861-7884.
18. Bhargava VP, Ko E, Willems M, Mercola S. Quantitative transcriptomics using designed primer-based amplification. *Sci Rep* 2013; 3: 1740.
19. Casbon JA, Osborne RJ, Brenner S. A method for counting PCR template molecules with application to next-generation sequencing. *Nucl Acids Res* 2011; 39: 81.
20. Casbon JA, Osborne RJ, Brenner S. A method for counting PCR template molecules with application to next-generation sequencing. *Nucl Acids Res* 2011; 39: e81.
21. Bhargava VH, Head SR, Ordoukhanian P. Technical variations in low-input RNA-seq methodologies. *Sci Rep* 2014; 4: 3678.
22. Archer J, Baillie G, Watson SJ. Analysis of high-depth sequence data for studying viral diversity: A comparison of next generation sequencing platforms using Segminator II. *BMC Bioinform* 2012; 13: 13-47.
23. Knierim EB, Lucke JM, Schwarz M, Schuelke D. Systematic comparison of three methods for fragmentation of long-range PCR products for next generation sequencing. *PLoS One* 2011; 6: 28240.
24. Steven R, Head H, Komori Kiyomi. Library construction for next-generation sequencing: Overviews and challenges. *BioTech* 2014; 56: 61-77.
25. Erwin L, Dijkvan YJ, Claude T. Library preparation methods for next-generation sequencing: Tone down the bias. *Expe Cell Res* 2014; 12-20.
26. Steven R, Head H, Komori K. Library construction for next-generation sequencing: Overviews and challenges. *BioTech* 2014; 56: 61-77.
27. Kozarewa IZ, Ning MA, Quail MJ. Amplification-free illumina sequencing-library preparation facilitates improved mapping and assembly of (G+C)-biased genomes. *Turn Nat Meth* 2009; 6: 291-295.
28. Marine R, Polson SW, Ravel J. Evaluation of a transposase protocol for rapid generation of shotgun high-throughput sequencing libraries from nanogram quantities of DNA. *Appl Environ Micro* 2011; 77: 8071-8079.
29. Aird D, Ross MG, Chen WS. Analyzing and minimizing PCR amplification bias in Illumina sequencing libraries. *GNIR Geno Bio* 2011; 12: 18.
30. Bhargava VH, Head SR, Ordoukhanian P. Technical variations in low-input RNA-seq methodologies. *Sci Rep* 2014; 4: 3678.
31. Seguin Orlando A, Schubert M, Clary J. Ligation bias in illumina next-generation DNA libraries: implications for sequencing ancient genomes. *PLoS One* 2013; 8: 78575.
32. Dabney J, Meyer M. Length and GC-biases during sequencing library amplification: a comparison of various polymerase-buffer systems with ancient and modern DNA sequencing libraries. *BioTech* 2012; 52: 87-94.
33. Oyola SO, Otto TD, Gu Y. Optimizing Illumina next-generation sequencing library preparation for extremely AT-biased genomes. *BMC Genomics* 2012; 13: 1.
34. Hayashi S, Murakami S, Omagari K. Characterization of novel entecavir resistance mutations. *J Hepat* 2015; 52: 87-94.
35. Syed F, Grunenwald H, Caruccio N. Next-generation sequencing library preparation: Simultaneous fragmentation and tagging using in vitro transposition. *Nat Met Appl Note* 2009; 6.
36. Parkinson NJ, Maslau SB, Ferney H. Preparation of high-quality next-generation sequencing libraries from picogram quantities of target DNA. *Genome Res* 2012; 22: 125-133.
37. Zhuang F, Fuchs R T, Sun Z, Zheng Y. Structural bias in T4 RNA ligase-mediated 3'-adapter ligation. *Robb Nucleic Acids Res* 2012; 40: 54.
38. Sorefan K, Pais H, Hall AE, Kozomara A. Reducing ligation bias of small RNAs in libraries for next generation sequencing. *Silence* 2012; 3: 4.
39. Oyola SO, Otto TD, Gu YG. Optimizing illumina next-generation sequencing library preparation for extremely AT-biased genomes. *BMC Genom* 2012; 13: 1.
40. Adey A, Morrison HG, Xun X. Rapid, low-input, low-bias construction of shotgun fragment libraries by high-density in vitro transposition. *Geno Biol* 2010; 11: 119.
41. Syed F, Grunenwald H, Caruccio N. Optimized library preparation method for next-generation sequencing. *Nat Met Appl Note* 2009.
42. Adey A, Morrison HG, Asan, Xun X, Kitzman JO. Rapid, low-input, low-bias construction of shotgun fragment libraries by high-density in vitro transposition. *Genome Biol* 2010; 11.

43. Grunenwald H, Baas B, Caruccio NF. Syed Rapid, high-throughput library preparation for next-generation sequencing. *Nat Met* 2010; 7: 1548-7091.
44. Steven R, Head H, Komori K. Library construction for next-generation sequencing: Overviews and challenges. *BioTech* 2014; 56: 61-77.
45. Lennon NJ, Lintner RE, Anderson S. A scalable, fully automated process for construction of sequence-ready barcoded libraries for 454. *Genome Biol* 2010; 11.
46. Sittka A, Lucchini S, Papenfort K. Deep sequencing analysis of small noncoding RNA and mRNA targets of the global post-transcriptional regulator. *PLoS Gene* 2008; 4: 1000163.
47. Erwin L, Dijkvan YJ, Claude T. Library preparation methods for next-generation sequencing: Tone down the bias. *Expe Cell Res* 2014; 12-20.
48. Kozarewa IZ, Ning MA, Quail MJ. Amplification-free illumina sequencing-library preparation facilitates improved mapping and assembly of (G+C)-biased genomes. *Turn Nat Meth* 2009; 6: 291-295.
49. Marine R, Polson SW, Ravel J. Evaluation of a transposase protocol for rapid generation of shotgun high-throughput sequencing libraries from nanogram quantities of DNA. *Appl Environ Micro* 2011; 77: 8071-8079.
50. Aird D, Ross M G, Chen WS. Analyzing and minimizing PCR amplification bias in Illumina sequencing libraries. *GNIR Geno Bio* 2011; 12: 18.
51. Seguin Orlando A, Schubert M, Clary J. Ligation bias in illumina next-generation DNA libraries: implications for sequencing ancient genomes. *PLoS One* 2013; 8: 78575.
52. Dabney J, Meyer M. Length and GC-biases during sequencing library amplification: a comparison of various polymerase-buffer systems with ancient and modern DNA sequencing libraries. *Biotech* 2012; 52: 87-94.
53. Oyola SO, Otto TD, Gu Y. Optimizing illumina next-generation sequencing library preparation for extremely AT-biased genomes. *BMC Genomics* 2012; 13: 1.
54. Parkinson NJ, Maslau SB, Ferney H. Preparation of high-quality next-generation sequencing libraries from picogram quantities of target DNA. *Genome Res* 2012; 22: 125 -133.
55. Zhuang F, Fuchs RT, Sun Z, Zheng Y. Structural bias in T4 RNA ligase-mediated 3'-adapter ligation. *Robb Nucleic Acids Res* 2012; 40: 54.
56. Adey A, Morrison H G, Xun X. Rapid, low-input, low-bias construction of shotgun fragment libraries by high-density in vitro transposition. *Geno Biol* 2010; 11: 119.
57. Syed F, Grunenwald H, Caruccio N. Next-generation sequencing library preparation: Simultaneous fragmentation and tagging using in vitro transposition. *Nat Met Appl Note* 2009; 6.
58. Syed F, Grunenwald H, Caruccio N. Optimized library preparation method for next-generation sequencing. *Nat Met Appl Note* 2009.
59. Bose ME, He J, Shrivastava S, Nelson MI. Sequencing and analysis of globally obtained human respiratory syncytial virus A and B genomes. *PLoS One* 2015; 10: 0120098.
60. Thaitrong N, Kim H, Renzi RF. Quality control of next-generation sequencing library through an integrative digital microfluidic platform. *Electroph* 2012; 33: 3506-3513.
61. Lennon NJ, Lintner RE, Anderson S. A scalable, fully automated process for construction of sequence-ready barcoded libraries for 454. *Genome Biol* 2010; 11.
62. Zhang JC, Badr A, Zhang G. The impact of next-generation sequencing on genomics. *J Genet Genomics* 2011; 38: 95-109.
63. Prosperi MCF, Prosperi L, Bruselles A. Combinatorial analysis and algorithms for quasispecies reconstruction using next-generation sequencing. *BMC Bio* 2011; 12: 5.
64. Dohm JC, Lottaz C, Borodina T, Himmelbauer H. Substantial biases in ultra-short read datasets from high-throughput DNA sequencing. *Nucl Acids Res* 2008; 36: 105.
65. Prosperi MCF, Salemi M. Software for viral quasispecies reconstruction from next-generation sequencing data. *Bioinformatics* 2012; 28: 132-133.
66. Pybus OG, Rambaut A. Evolutionary analysis of the dynamics of viral infectious disease. *Nat Rev Genet* 2009; 10: 540-550.
67. Szittya G, Moxon S, Pantaleo V. Dalmay structural and functional analysis of viral siRNAs. *PLoS Pathog* 2010; 6: 1000838.
68. Pabinger S, Dander A, Fischer M. A survey of tools for variant analysis of next-generation genome sequencing data. *Brief Bioinform* 2013.
69. Willerth SM, Pedro HA, Pachter L. Development of a low bias method for characterizing viral populations using next generation sequencing technology. *PLoS One* 2010; 5: 13564.
70. Pareek CS, Smoczynski R. Sequencing technologies and genome sequencing. *J App Genet* 2011; 52: 413-435.
71. Hendrix D, Levine M, Shi W. miRTRAP, a computational method for the systematic identification of miRNAs from high throughput sequencing data. *Geno Biol* 2010; 11: 39.
72. Miller R, Koren S, Sutton G. Assembly algorithms for next-generation sequencing data. *Genomics* 2010; 95: 315-327.
73. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 2009; 25: 1754-1760.
74. Marioni JC, Mason CE, Mane SM. RNA-seq: an assessment of technical reproducibility and comparison with gene expression arrays. *Genome Res* 2011; 8: 1509-1517.
75. Rosperi MCF, Salemi M. Software for viral quasispecies reconstruction from next-generation sequencing data. *Bioinformatics* 2012; 28: 132-133.
76. Szittya G, Moxon S, Pantaleo V. Dalmay Structural and functional analysis of viral siRNAs. *PLoS Pathog* 2010; 6: 1000838.
77. Szittya G, Moxon S, Pantaleo V. Dalmay Structural and functional analysis of viral siRNAs. *PLoS Pathog* 2010; 6: e1000838

78. Pabinger S, Dander A, Fischer M. A survey of tools for variant analysis of next-generation genome sequencing data. *Brief Bioinform* 2013.
79. Willerth SM, Pedro HA, Pachter L. Development of a low bias method for characterizing viral populations using next generation sequencing technology. *PLoS One* 2010; 5: 13564.
80. Pareek CS, Smoczynski R. Sequencing technologies and genome sequencing. *J App Genet* 2011; 52: 413-435.
81. Hendrix D, Levine M, Shi W. miRTRAP, a computational method for the systematic identification of miRNAs from high throughput sequencing data. *Geno Biol* 2010; 11: 39.
82. Miller R, Koren S, Sutton G. Assembly algorithms for next-generation sequencing data. *Genomics* 2010; 95: 315-327.
83. Li H., Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 2009; 25: 1754-1760.
84. Odovar N, Goic B, Blanc H. In silico reconstruction of viral genomes from small RNAs improves virus-derived small interfering RNA profiling. *J Virol* 2011; 85: 11016-11021.
85. Roberts A, Pimentel H, Trapnell C. Identification of novel transcripts in annotated genomes using RNA-Seq. *Bioinformatics* 2011; 27: 2325-2329.
86. Larson DE, Harris CC, Chen K. Somatic Sniper: identification of somatic point mutations in whole genome sequencing data. *Bioinformatics* 2012; 28: 311-317.
87. Zywicki M, Bakowska-Zywicka K, Polacek N. Revealing stable processing products from ribosome-associated small RNAs by deep-sequencing data analysis. *Nucl Acids Res* 2012; 40: 4013-4024.
88. Mathelier A, Carbone A. mi-RNA: finding microRNAs with high accuracy and no learning at genome scale and from deep sequencing data. *Bioinformatics* 2010; 26: 2226-2234.
89. Hendrix D, Levine M, Shi W. miRTRAP, a computational method for the systematic identification of miRNAs from high throughput sequencing data. *Genome Biol* 2010; 1: 39.
90. Liu CM, Wong T, Wu E. SOAP3: ultra-fast GPU-based parallel alignment tool for short reads. *Bioinformatics* 2012; 28: 878-879.
91. Bimber BN, Dudley DM, Lauck M. Whole-genome characterization of human and simian immunodeficiency virus intrahost diversity by ultradeep pyrosequencing. *Maffitt J Virol* 2010; 84: 12087-12092.
92. Mokili JL, Rohwer F, Dutilh BE. Metagenomics and future perspectives in virus discovery. *Curr Opin Virol* 2012; 2: 63-77.
93. MacLean DJ, Jones DJ. Studholme. Application of next-generation sequencing technologies to microbial genetics. *Nat Rev Microbial* 2009; 7: 287-296.
94. Bexfield N, Kellam P. Metagenomics and the molecular identification of novel viruses. *Vet J* 2011; 190: 191-198.
95. Nakamura S, Yang CS, Sakon N. Direct metagenomic detection of viral pathogens in nasal and fecal specimens using an unbiased high-throughput sequencing approach. *PLoS One* 2009; 42: 19.
96. Nakamura S, Yang CS, Sakon N. Direct metagenomic detection of viral pathogens in nasal and fecal specimens using an unbiased high-throughput sequencing approach. *PLoS One* 2009; 4: 4219.
97. Li L, Delwart E. From orphan virus to pathogen: the path to the clinical lab. *Curr Opin Virol* 2011; 1: 282-288.
98. Rosario K, Breitbart M. Exploring the viral world through metagenomics. *Curr Opin Virol* 2011; 1: 289-297.
99. Duffy S, Shackelton LA, Holmes EC. Rates of evolutionary change in viruses: patterns and determinants. *Nat Rev Gene* 2008; 9: 267-276.
100. Beerenwinkel N, Zagordi O. Ultra-deep sequencing for the analysis of viral populations. *Curr Opin Virol* 2011; 413-418.
101. Kuroda M, Katano H, Nakajima N. Characterization of quasispecies of pandemic 2009 influenza A virus (A/H1N1/2009) by de novo sequencing using a next-generation DNA sequencers. *PLoS One* 2010; 5.
102. Qiu P, Stevens R, Wei B, Lahser F. HCV genotyping from NGS short reads and its application in genotype detection from HCV mixed infected plasma. *PLoS One* 2015; 10: 0122082.
103. Nakamura S, Yang CS, Sakon N. Direct metagenomic detection of viral pathogens in nasal and fecal specimens using an unbiased high-throughput sequencing approach. *PLoS One* 2009; 42: 1.

#### \*Correspondence to

Desh Deepak Singh

Amity institute of Biotechnology

Amity University

India