Review

RNA Sequencing and its Applications in Cancer Diagnosis and Targeted Therapy

Mimi Wan, MD, PhD;* Jianhui Wang, PhD; Xiaobin Gao, MPh; Jeffery Sklar, MD, PhD

Department of Pathology, Yale University School of Medicine, New Haven, CT

High throughput DNA and RNA sequencing (DNA-Seq and RNA-Seq) is increasingly impacting the clinical practice of medicine. RNA-Seq has so far had a smaller role in the clinical practice, but it has advantageous features and is complementary to DNA-Seq. RNA-Seq can profile the abundance and composition of the entire transcriptome, including both mRNA and non-coding RNA. It is thus capable of revealing diverse functional and structural changes affecting genes, such as gene overexpression, silencing and various abnormalities and alterations among which may be substitutions, deletions, inversions, alternative splicing and gene fusions. As cancers are characterized by many of these changes, RNA-Seq can be valuable for diagnosing and characterizing tumors. Here we will describe the use of RNA-Seq in cancer diagnosis and personalized therapy, with an emphasis on the detection of fusion transcripts, which are frequently associated with cancer and are often drug targets for cancer therapy.

[N A J Med Sci. 2014;7(4):156-162. **DOI:** *10.7156/najms.2014.0704156]*

Key Words: *RNA-Seq, FFPE tissue, whole transcriptome RNA-Seq, targeted RNA-Seq, precision medicine, personalized medicine, targeted therapy*

INTRODUCTION

Over the past decade, the next generation sequencing (NGS) technologies for rapid, high-throughput analysis of the genome and transcriptome have revolutionized basic biomedical research and increasingly found a place in diagnostic medicine.[1,2] DNA-Seq, including targeted and whole-exome sequencing, has been highly successful for detecting germline and somatic mutations. However, DNA-Seq cannot detect gene rearrangements unless expensive whole-genome sequencing is performed, and several other crucial assessments are beyond the reach of DNA-Seq, including gene activity and alternative splicing of RNA. These limitations of DNA-Seq are addressable using RNA-Seq, which can detect gene rearrangement, RNA abundance and splicing as well as sequence variations (such as small mutations in expressed regions of the genome). As DNA-Seq detects genetic lesions while general RNA-Seq reveals their consequences, the two are complementary. This review is intended as a primer for the RNA-Seq technology and its application in cancer diagnosis and therapy. For in-depth discussions of the technology, readers are referred to several recent reviews.[3-5]

OVERVIEW OF RNA-SEQ TECHNOLOGY

The basic components of the RNA-Seq technology in the clinical context include five major steps: RNA extraction from clinical samples, NGS library preparation, sequencing, data analysis, and data interpretation/reporting, as depicted in **Figure 1** and detailed below.

RNA extraction. Body fluids, fresh-frozen or formalin-fixed, paraffin-embedded (FFPE) tissues are often the sources of RNA. RNA from body fluids and frozen tissues can be purified using routine RNA extraction methods or commercial kits, such as the RNeasy kit (Qiagen). For the FFPE tissues, cancer cells are excised from tissue sections usually by manual microdissection or by laser capture microdissection. RNA is then isolated using kits such as the RNeasy FFPE kit (QIAGEN), FormaPure (Agencourt), and the Human FFPE RNA-Seq Multiplex Systems (Ovation). For RNA from body fluids and frozen tissues, 10 ng can be sufficient for RNA-Seq, but > 100ng RNA is often needed if isolated from FFPE tissues, due to its fragmented nature and generally inferior quality.

NGS library preparation. NGS libraries consist of cDNAs converted from the RNA and end-adapted for sequencing. Libraries for whole transcriptome are prepared using kits that are available from several vendors and fit the respective sequencing platforms, such as Illumina's TruSeq Stranded Total RNA Sample Prep Kit and Ion Torrent's Ion Total RNA-Seq Kit. Kits can also be purchased to generate NGS libraries for targeted analysis of candidate RNAs. For example, Illumina's TruSeq Targeted RNA-Expression Kit offers customizable mid- to high-plex gene expression profiling of genes involved in specific pathways and disease states. Furthermore, Enzymatics's Archer FusionPlex Assays,

Received: 09/16/2014; Revised: 09/30/2014; Accepted: 10/01/2014
***Corresponding Author:** Department of Pathology, Yale University School of Medicine, 310 Cedar Street, New Haven, CT 06520. Tel: 203-737-6061. Fax: 203-785-3896.
(Email: mimi.wan@yale.edu)

based on "anchored multiplex PCR",[6] can detect gene fusions and mutations in different cancers without prior knowledge of the specific fusion partners.

Sequencing. The NGS principles and different platforms have been covered in numerous reviews.[1,7-9] A comparison of different platforms is listed in **Table 1**. Illumina and Ion Torrent platforms are commonly used in clinical laboratories. Illumina's Nextseq 500 and Ion Proton are suitable for sequencing the whole transciptome, whereas Illumina's MiSeq (MiSeqDx) and Ion Torrent's PGM are ideal for targeted sequencing.

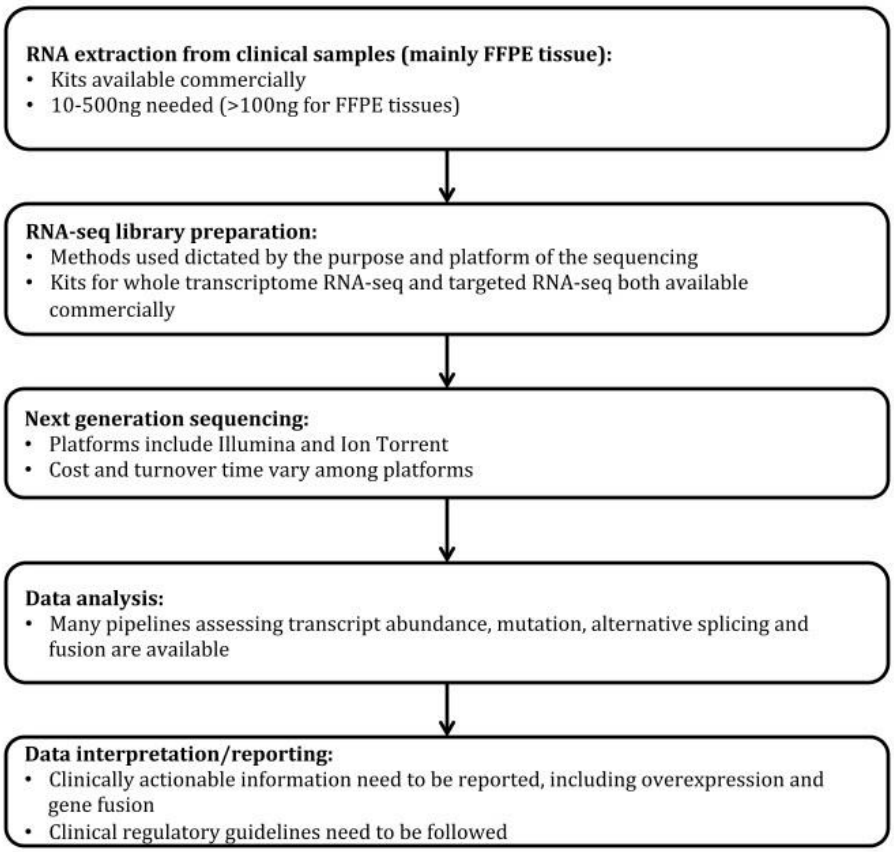

Figure 1. A flow chart of RNA-Seq in clinical settings.

Table 1. A comparison of commonly used deep sequencing platforms.

Company	Platform	Seqencing Mechanism	Current Model	Maximum Read Length (bp)	Highest Throughput	Turnaround Time	Error Rate (%)
Roche	454	Synthesis reaction + pyrophosphate release, chemiluminescence detection	GS FLX+	1000	700Mb	23 hours	1
			GS Junior	400	35Mb	10 hours	1
Illumina	Solexa	Synthesis reaction + colorimetric label detection	HiSeqX	150	1.8Tb	3 days	0.4
			HiSeq 2500	150	180Gb	40 hours	0.4
			NextSeq 500	150	39Gb	26 hours	0.4
			MiSeq (MiSeqDx)	300	15Gb	8.5 hours	0.4
Life Technology	SOLiD	Ligation reaction + colorimetric label detection	5500xl W	50	320Gb	6 days	0.1
	Ion Torrent	Synthesis reaction + proton release and pH sensing	PGM	400	2Gb	3 hours	1
			Proton	400	10Gb	4 hours	1
Oxford Nanopore	Nanopore	None reaction, single molecule electronics-based sensing	GridION	100,000	185Mb	6 hours	4
			MinION	100,000	28G/10nodes. 1day	N/A*	4
Pacific Biosciences	PacBio	Synthesis reaction + single molecule real-time (SMRT) colorimetric label detection	PacBio RS II	30,000	350Mb	10 hours	15

* GridION has no fixed turnaround time because it is designed to be a cluster aggregated with nodes. A user can run one or more nodes for minutes or days according to how much data is needed to complete the experiment.

Data Analysis. The raw sequence reads generated by the sequencing machines are analyzed using various bioinformatics tools. The reads are first aligned to reference genomes to construct transcriptomes. Transcript abundance is calculated using read counts, while alternative splicing and gene mutations are inferred by comparing the transcriptomes with the reference genomes.

Commonly used bioinformatics tools are classified based on their functions (columns 1-2, **Table 2**). Different programs in the same functional categories do not always produce consistent results even though they may use the same algorithms (column 4), and so the data must be interpreted with caution. These software programs are mostly (69 of 72) from open-source communities and thus freely available; indeed, such open source software (OSS) programs have been a cornerstone in the RNA-Seq technology.[10] The 69 OSS are mainly licensed by Artistic, BSD, GPL or LGPL,[11] with > 50% (35 of 69) by GPL according to which new tools using the 35 softwares are obligated to be similarly open-sourced. In clinical applications, the OSS codes often require modifications to become user-friendly and need-tailored, and any fault in the operation must be quickly corrected, which requires substantial bioinformatics expertise. Alternatively, clinical labs may resort to commercial software programs, which are typically more user-friendly and well supported, though such programs are costly and inflexible and thus generally less useful than OSS once a user becomes familiar with their operation.[12] The bioinformatics tools listed in the table are relatively new, developed since 2008 (column 6, **Table 2**),[13-81] and additional software packages are expected to emerge in parallel with improved or novel sequencing platforms.

Table 2. A list of bioinformatics tools for RNA-Seq data analysis.

Function	Sub-Category	Tool Name	Algorithm Note	Code License	Published Year and Reference
Alignment	Non Spliced	Maq	spaced seed	GPL	2008[13]
		SOAP	seed and hash look-up table	GPL	2008[14]
		SeqMap	spaced seed	codes open to academia	2008[15]
		RazerS	spaced seed/q-gram	GPL	2009[16]
		Bowtie	Burrows-Wheeler transform	Artistic and GPL	2009[17]
		BWA	Burrows-Wheeler transform/ Smith-Waterman algorithm	GPL	2009[18]
		BFAST	Smith-Waterman algorithm	GPL	2009[19]
		SHRiMP	Smith-Waterman algorithm/q-gram	BSD	2009[20]
		NovoAlign	Needleman–Wunsch algorithm	codes not open but binaries free for academic	2009[21]
		SOAP2	Needleman–Wunsch algorithm	GPL	2010[22]
		GNUMAP	Needleman–Wunsch algorithm	codes open	2010[23]
		Stampy	Burrows-Wheeler transform	Codes open to academic	2011[24]
		SeqAlto	Needleman–Wunsch algorithm	codes not open but binaries free for academic	2012[25]
		Mosaik	Smith-Waterman algorithm	GPL	2014[26]
	Spliced	QPALMA	large margin algorithm	GPL	2008[27]
		TopHat	candidate exons pairing , implanted bowtie	codes open	2009[28]
		SMALT	Smith-Waterman algorithm	GPL	2010[29]
		GSNAP	SNP-tolerant	codes open	2010[30]
		PALMapper	machine learning algorithm, implanted QPALMA	GPL	2010[31]
		SplitSeek	anchors extension	GPL	2010[32]
		SpliceMap	Optional filtering	codes open	2010[33]
		MapSplice	Anchoring exons flanking alignment	GPL	2010[34]
		HMMsplicer	Hidden Markov Model	codes open	2010[35]
		STAR	Maximal Mappable Prefix	GPL	2012[36]
		GEM	Filtration/BWT	codes open to academic	2012[37]
		Subread	seed-and-vote	GPL	2013[38]
Expression Differences Analysis	Non-Bioconductor	Useq	negative binomial distribution	BSD	2008[39]
		Cufflinks	Beta negative binomial distribution	BSL, codes open	2009[40]
		TSPM	two-stage Poisson model	codes open	2011[41]
		RSEM	Dirichlet prior distribution	GPL	2011[42]
		NBPseq	Negative binomial distribution	GPL	2011[43]
		Samseq	Nonparametric method	LGPL	2011[44]
		BBseq	Negative binomial distribution	GPL	2011[45]
		Gfold	posterior distribution, single biological duplicate	codes open	2012[46]
	Bioconductor	Limma	voom transformation of counts	GPL	2005[47]
		Deseq	Negative binomial distribution	GPL	2010[48]
		baySeq	Negative binomial distribution	GPL	2010[49]
		NOIseq	Nonparametric method	Artistic	2011[50]
		edgeR	Negative binomial distribution	LGPL	2012[51]
		sSeq	Negative binomial distribution	GPL	2013[52]
		EBSeq	Negative binomial distribution	Artistic	2014[53]
Transcriptome Assembling		Trans-Abyss	mutiple k-mer assemblies	codes open to academic	2010[54]
		Trinity	single k-mer assemblies, dynamic filters	BSD	2011[55]
		Oases	mutiple k-mer assemblies, dynamic filters	GPL	2012[56]
		SOAPdenovo-Trans	mutiple k-mer assemblies, sparse-pregraph	GPL	2014[57]
Mutation Detection		SNPiR	implanted GATK	codes open	2013[58]
		MMAPPR	implanted Samtools pileup	codes open to academic	2013[59]
		Rnaseqmut	Normal-Tumor comparison	codes open	2013[60]

Table 2. A list of bioinformatics tools for RNA-Seq data analysis. (Continued)

Function	Sub-Category	Tool Name	Algorithm Note	Code License	Published Year and Reference
Alternative Splicing Detection		MISO	Markov Chain Monte Carlo	BSD	2010[61]
		ALEXA-Seq	Identity of the subset of differentially expressed features	GPL	2010[62]
		Alt Event Finder	identity novel cassette exon event	codes open	2012[63]
		DEXseq	generalized linear models	GPL	2012[64]
		r-Diff	with non-parametric test	GPL	2013[65]
		SAJR	Binomial Generalized Linear Model	codes open	2013[66]
		ARH-seq	entropy, Weibull distribution	codes open	2014[67]
		RSVP	ORF graph	codes open	2014[68]
		SwitchSeq	identity of transcript changes across conditions	GPL	2014[69]
Fusion Detection	Single End Reads	GSNAP	SNP-tolerant	codes open	2010[30]
		SplitSeek	anchors extension	GPL	2010[32]
		Tophat_fusion	anchors extension	codes open	2011[70]
		FusionMap	Making pseudo PE reads	codes not open but binaries free for academic	2011[71]
		FusionFinder	Making pseudo PE reads	GPL	2012[72]
		GSTRUCT	probabilistic models	codes open	2012[73]
	Pair End Reads	FusionSeq	building fusion junction library	codes open	2010[74]
		Fusionhunter	with some Kent Source Code	GPL	2011[75]
		Chimerascan	realignning the trimmed reads	GPL	2011[76]
		snowshoes-FTD	with prediction of fusion mechanism	GPL	2011[77]
		DeFuse	confidence measure	codes open	2011[78]
		ShortFuse	estimating fusion transcript aboundances	codes open	2011[79]
		EricScript	recalibrating junction reference	GPL	2012[80]
		SOAPfuse	can detect low coverage	codes open	2013[81]
		STAR	Maximal Mappable Prefix	GPL	2013[36]

Data interpretation and reporting. Data are evaluated for clinical relevance and significance based on its nature and published literature. For whole transcriptome analysis, it can be challenging to interpret the data rapidly and accurately for clinical use. For targeted RNA-Seq, commercial companies developing specific RNA-Seq panels usually also develop special software pipelines that can generate reportable results from raw data. Unlike simple pathogenic DNA variants such as single nucleotide variants (SNVs) and insertions/deletions (INDELs), information about gene expression and fusions is not as well represented in databases such as Catalogue of Somatic Mutations in Cancer (COSMIC), the Online Mendelian Inheritance in Man (OMIM), and the Human Gene Mutation Database (HGMD). NCBI UniGene will be helpful in some extent as a source for gene expression information with regard to specific genes. National Cancer Institute's Cancer Genome Anatomy Project (CGAP) website is a good resource for gene expression and fusion in cancer (http://cgap.nci.nih.gov/Catalog). As part of CGAP, the Mitelman Database of Chromosome Aberrations and Gene Fusions in Cancer has some very useful tools such as "Clinical Associations Searcher".

APPLICATION OF RNA-SEQ TO CANCER DIAGNOSIS AND TARGETED THERAPY

Cancer is characterized by a dazzling array of genetic lesions directly affecting genes, including point mutation, insertion, deletion, translocation, exon-skipping and gene fusion. If the mutant genes are transcribed, these lesions become detectable by RNA-Seq. At the same time, RNA-Seq measures the transcript abundance, thus revealing the expression levels of the mutant genes.[82,83] RNA-Seq can thus identify biomarkers for cancer risks, subtypes and stages of progression, providing crucial insights into cancer diagnosis, prognosis and potential personalized therapies. For example, whole transcriptome RNA-Seq reveals that T-cell acute lymphoblastic leukemia (T-ALL) have diverse defects including point mutations, insertions, deletions, translocations, exon-skipping and gene fusion to novel partners such as kinases.[83] Based on the known and novel driver mutations at protein coding genes and their mRNA expression levels, the patients are classified into different T-ALL subtypes, which has important implication for personalized medicine. RNA-Seq similarly helps identify biomarkers for breast cancers. Specifically, RT-PCR analysis of FFPE tumor samples from a cohort of 136 breast cancer patients has previously revealed 21 genes the expression of which is associated with breast cancer recurrence (the Oncotype DX Breast Cancer Assay).[84] RNA-Seq of the same tumor samples confirmed the RT-PCR data, and uncovered more than two thousand additional RNAs that are also strongly associated with breast cancer recurrence risk.[85] RNA-Seq is particularly useful for detecting gene fusions, as described below.

RNA-SEQ DETECTION OF GENE FUSIONS IN CANCER

A hallmark of cancer is the fusion of segments of one gene to that of another, which can result from multiple types of genetic lesions, including translocation, deletion and inversion. Depending on the fusion partners, gene fusion can cause aberrant activation of oncogenes. Consequently, gene fusion analysis is widely used for cancer diagnosis (i.e, RET-PTC for thyroid cancer), prognosis (i.e, TMPRSS2-ERG for prostate cancer) and targeted therapy (i.e, EML4-ALK for lung cancer). Gene fusions are usually detected using FISH or RT-PCR. However, FISH is technically challenging and results are sometimes equivocal, while conventional RT-PCR requires prior knowledge of both fusion partners, which can be problematic because very often, a gene is fused to multiple possible partners and each partner may have multiple variants, making it impractical to interrogate for all possible fusion events.

RNA-Seq provides a powerful solution for detecting all fusion events, as long as the transcripts are expressed. Both whole transcriptome and targeted sequencing have been used. For example, whole transcriptome sequencing reveals over one hundred fusion events in three independent studies, each involving over 70 patients.[85-87] Interestingly, the high

frequency of fusion transcripts are correlated with poor outcome in human breast cancer patients, suggesting the prognostic value of the fusion transcripts.

Major drawbacks associated with whole transcriptome sequencing are that it requires sophisticated bioinformatics analysis and is costly. In contrast, targeted sequencing involves relatively simple bioinformatics, and also gives greater depth of coverage at the same numbers of reads, resulting in higher sensitivity, specificity and cost-effectiveness; it may be the method of choice when looking for the fusion partners of a particular gene of interest (GOI). In targeted sequencing, the fusion transcripts to be sequenced are first selectively amplified by PCR-based method. A robust method, called Anchored Multiplex PCR (AMP), which is similar to conventional Rapid Amplification of cDNA Ends (RACE) technique, has been developed to amplify the fusion transcripts of GOI without prior knowledge of the fusion partners.[6] In AMP, RNA is first transcribed into cDNA, which is then ligated to an adaptor. The adaptor serves as an anchor for PCR that involves a universal primer annealing to the adaptor and a gene-specific primer targeting the GOI. The PCR product is then sequenced by Ion Torrent or Illumina platforms, which reveals all fusion partners of the GOI. Indeed, using AMP, Zheng, et al., detected not only known fusion targets such as ALK, RET and ROS1 in FFPE tissues but also novel gene fusions.[6] The sensitivity and specificity of targeted sequencing are superior to that achieved by standard FISH assays. The AMP technology is licensed to Enzymatics, and commercial kits called Archer FusionPlex Assays are available to detect gene fusions in lung cancer, hematological malignancies and sarcomas. (http://www.enzymatics.com/archer/fusionplex-assays/)

DEVELOPMENT OF CLINICAL RNA-SEQ TESTS: PRACTICE GUIDANCE

In the light of changing FDA policies regarding laboratory developed tests (LDTs), special efforts should be made to understand the current regulatory environment. Before the new FDA policies regarding LDTs become officially active, it is recommended to follow the practice guidelines published by organizations such as American College of Medical Genetics (ACMG),[88] the US Centers for Disease Control and Preventions (CDC),[89] and the Association of Molecular Pathology (AMP).[90] As in the case of other clinical tests, for RNA-Seq, a number of factors need to be taken into consideration, including clinical utility, test validation, quality control, assay sensitivity and specificity, turnaround time, strategies for proficiency testing, and reference materials. Of note, RNA is highly labile and care must be taken to avoid RNase contamination. Additionally, RNA samples from clinical specimens such as FFPE tissues are often fragmented and require special chemistry for purification and constructing libraries. Thus, commercial kits are recommended for extracting RNA and making NGS libraries. Also, we recommend starting with a focused RNA-Seq panel, although in the future, when sequencing costs drop and bioinformatics tools are simplified, whole transcriptome analysis is likely to dominate in clinical labs.

CONCLUSIONS AND FUTURE PERSPECTIVES

RNA-Seq has proven invaluable for detecting aberrations in transcript abundance and/or structure in cancer, but there are multiple challenges and opportunities ahead.

First, the current major sequencing platforms (Illumina and Ion Torrent) only produce short sequencing reads (up to a few hundreds). Such short sequences can be difficult to assign to the genome in the cases of gene fusion or alternative splicing, causing ambiguities or errors in data interpretation. Longer sequencing reads (up to thousands) are possible using machines from certain manufacturers (e.g. Pacific Biosciences), but such "third generation sequencers" have relatively high error rates and are thus not commonly used in clinical settings. However, they would be extremely valuable once the error rates are improved.

Second, RNA-Seq is typically done at the cell population level. However, tumor tissues are heterogeneous, with individual cells within the same cancer possessing divergent transcriptomes. Such heterogeneity may be functionally relevant. In particular, cancer stem cells behave very differently from their progenies, and are crucial therapeutic targets. RNA-Seq at the single cell level is now feasible,[91] which can provide critical insights into cancer biology. However, to avoid sampling errors, many individual cells must be sequenced, and this is impractical in the clinical setting. Recently, a bar-code based, multiplex sequencing method has been developed that can interrogate thousand single cells at once in the same lane,[92] which is expected to find its way into clinical labs.

Finally, beside mRNA, non-coding RNA such as microRNA and lncRNA are also key players in cancer development and progression.[93-95] However, they have received less attention in the clinical labs - particulary so for LncRNA. LncRNA may therefore offer a very fertile ground for future discoveries.

ACKNOWLEDGMENTS

The authors thank Drs. Tian Chi, Pei Hui and Keji Zhao for reading the manuscript and for valuable comments.

CONFLICT OF INTEREST

None.

REFERENCES

1. Koboldt DC, Steinberg KM, Larson DE, Wilson RK, Mardis ER. The next-generation sequencing revolution and its impact on genomics. Cell. 2013;155(1):27-38.
2. Rivera CM1, Ren B. Mapping human epigenomes. Cell. 2013;155(1):39-55.
3. Wang Z, Gerstein M, Snyder M. RNA-Seq: a revolutionary tool for transcriptomics. Nat Rev Genet. 2009;10(1):57-63.
4. Martin JA1, Wang Z. Next-generation transcriptome assembly. Nat Rev Genet. 2011;12(10):671-682.
5. Ozsolak F, Milos PM. RNA sequencing: advances, challenges and opportunities. Nat Rev Genet. 2011;12(2):87-98.
6. Zheng Z, Liebers M, Zhelyazkova B, et al. Anchored multiplex PCR for targeted next-generation sequencing. Nat Med. [in press]
7. Metzker ML. Sequencing technologies - the next generation. Nat Rev Genet. 2010;11(1):31-46.
8. Mardis ER. A decade's perspective on DNA sequencing technology. Nature. 2011;470(7333):198-203.
9. Mardis ER. Next-generation sequencing platforms. Annu Rev Anal Chem. 2013;6:287-303.

10. Bonnal RJP, Aerts J, Githinji G, et al. Biogem: an effective tool-based approach for scaling up open source software development in bioinformatics. Bioinformatics. 2012;28(7):1035-1037.
11. St.Laurent AM. Understanding Open Source and Free Software Licensing. O'Reill Media Inc, Sebastopol, CA. 2004.
12. Vincent AT, Charette SJ. Freedom in bioinformatics. Front. Genet. 2014;5:259.
13. Li H, Ruan J, Durbin R. Mapping short DNA sequencing reads and calling variants using mapping quality scores. Genome Res. 2008; 18(11):1851-1858.
14. Li R, Li Y, Kristiansen K, et al. SOAP: short oligonucleotide alignment program. Bioinformatics. 2008;24(5):713-714.
15. Jiang H, Wong WH. SeqMap: mapping massive amount of oligonucleotides to the genome. Bioinformatics. 2008;24(20):2395-2396.
16. Weese D, Emde AK, Rausch T, et al. RazerS - fast read mapping with sensitivity control. Genome Res. 2009;19(9):1646-1654.
17. Langmead B, Trapnell C, Pop M, et al. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. 2009; 10(3):R25.
18. Li H, Durbin R. Fast and accurate short read alignment with Burrows–Wheeler transform. Bioinformatics. 2009;25(14):1754-1760.
19. Homer N, Merriman B, Nelson SF. BFAST: an alignment tool for large scale genome resequencing. PLoS One. 2009;4(11):e7767.
20. Rumble SM, Lacroute P, Dalca AV, et al. SHRiMP: Accurate Mapping of Short Color-space Reads. PLoS Comput Biol. 2009; 5(5):e1000386.
21. http://www.novocraft.com/main/page.php?s=novoalign.
22. Li R, Yu C, Li Y, et al. SOAP2: an improved ultrafast tool for short read alignment. Bioinformatics. 2009; 25(15):1966-1967.
23. Clement NL, Snell Q, Clement MJ, et al. The GNUMAP algorithm: unbiased probabilistic mapping of oligonucleotides from next-generation sequencing. Bioinformatics. 2010; 26(1):38-45.
24. Gerton L, Goodson M. Stampy: A statistical algorithm for sensitive and fast mapping of Illumina sequence reads. Genome Res. 2011; 21(6):936-939.
25. Mu JC, Jiang H, Kiani A, et al. Fast and accurate read alignment for resequencing. Bioinformatics. 2012;28(18):2366-2373.
26. Lee W-P, Stromberg MP, Ward A, et al. MOSAIK: A Hash-Based Algorithm for Accurate Next-Generation Sequencing Short-Read Mapping. PLoS ONE. 2014; 9(3):e90581.
27. Bona FD, Ossowski S, Schneeberger K, et al. Optimal spliced alignments of short sequence reads. Bioinformatics. 2008;24(16):i174-i180.
28. Trapnell C, Pathter L, Salzberg SL. TopHat: discovering splice junctions with RNA-Seq. Bioinformatics. 2009;25(9):1105-1111.
29. http://www.sanger.ac.uk/resources/software/smalt/.
30. Wu TD, Nacu S. Fast and SNP-tolerant detection of complex variants and splicing in short reads. Bioinformatics. 2010;26(7):873-881.
31. Jean G, Kahles A, Sreedharan VT, De Bona F, Rätsch G. RNA-Seq read alignments with PALMapper. Curr Protoc Bioinformatics. 2010; Chapter 11: Unit 11.6.
32. Ameur A, Wetterbom A, Feuk L, Gyllensten U. Global and unbiased detection of splice junctions from RNA-seq data. Genome Biol. 2010;11(3):R34.
33. Au KF, Jiang H, Lin L, Xing Y, Wong WH. Detection of splice junctions from paired-end RNA-seq data by SpliceMap. Nucl. Acids Res. 2010;38(14):4570-4578.
34. Wang K, Singh D, Zeng Z, et al. MapSplice: Accurate mapping of RNA-seq reads for splice junction discovery. Nucl Acids Res. 2010;38(18):e178.
35. Dimon MT, Sorber K, DeRisi JL. HMMSplicer: A Tool for Efficient and Sensitive Discovery of Known and Novel Splice Junctions in RNA-Seq Data. PLoS ONE. 2010;5(11):e13875
36. Dobin A, Davis C, Schlesinger F, et al. STAR: ultrafast universal RNA-seq aligner. Bioinformatics. 2013;29(1):15-21.
37. Marco-Sola S, Sammeth M, Guigó R, Ribeca P. The GEM mapper: fast, accurate and versatile alignment by filtration. Nat Methods. 2012;9(12):1185-1188.
38. Liao Y, Smyth GK, Shi W. The Subread aligner: fast, accurate and scalable read mapping by seed-and-vote. Nucleic Acids Res. 2013;41(10):e108.
39. Nix DA, Courdy SJ, Boucher KM. Empirical methods for controlling false positives and estimating confidence in ChIP-Seq peaks. BMC Bioinformatics. 2008;9:523.
40. Trapnell C, Roberts A, Goff L, et al. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. Nature Protoc. 2012;7(3):562-578.
41. Auer PL, Doerge R. A Two-Stage Poisson Model for Testing RNA-Seq Data. Statistical Applications in Genetics and Molecular Biology. 2011;10(1):1-26.
42. Li B, Dewey CN. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics. 2011;12:323.
43. Cumbie JS, Kimbrel JA, Di Y, et al. GENE-Counter: A Computational Pipeline for the Analysis of RNA-Seq Data for Gene Expression Differences. PLoS One. 2011;6(10):e25279.
44. Li J, Tibshirani R. Finding consistent patterns: A nonparametric approach for identifying differential expression in RNA-Seq data. Stat Methods Med Res. 2013;22(5):519-536.
45. Zhou YH, Xia K, Wright FA. A powerful and flexible approach to the analysis of RNA sequence count data. Bioinformatics. 2011;27(19):2672-2678.
46. Feng J, Meyer CA, Wang Q, Liu JS, Liu XS, Zhang Y. GFOLD: a generalized fold change for ranking differentially expressed genes from RNA-seq data. Bioinformatics. 2012;28(21):2782-2788.
47. Smyth GK. Limma: linear models for microarray data. In Gentleman R, Carey V, Dudoit S, Irizarry R and Huber W (eds.), Bioinformatics and Computational Biology Solutions Using R and Bioconductor. Springer, New York. 2005;pp.397-420.
48. Anders S, Huber W. Differential expression analysis for sequence count data. Genome Biol. 2010;11(10):R106.
49. Hardcastle TJ, Kelly KA. BaySeq: Empirical Bayesian methods for identifying differential expression in sequence count data. BMC Bioinformatics. 2010;11:422.
50. Tarazona S, Garcia-Alcalde F, Dopazo J, Ferrer A, Conesa A. Differential expression in RNA-seq: a matter of depth. Genome Res. 2011;21(12):2213-2223.
51. Robinson MD, McCarthy D, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics. 2010;26(1):139-140.
52. Yu D, Huber W, Vitek O. Shrinkage estimation of dispersion in Negative Binomial models for RNA-seq experiments with small sample size. Bioinformatics. 2013;29(10):1275-1282.
53. Leng N, Dawson JA, Kendziorski C. EBSeq: An R package for gene and isoform differential expression analysis of RNA-seq data. R package version 1.5.3, 2014. https://www.biostat.wisc.edu/~kendzior/EBSEQ/.
54. Robertson G, Schein J, Chiu R, et al. De novo assembly and analysis of RNA-seq data. Nat Methods. 2010;7(11):909-912.
55. Grabherr MG, Haas BJ, Yassour M, et al. Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat Biotechnol. 2011;29(7):644-652.
56. Schulz MH, Zerbino DR, Vingron M, Birney E. Oases: robust de novo RNA-seq assembly across the dynamic range of expression levels. Bioinformatics. 2012;28(8):1086-1092.
57. Xie Y, Wu G, Tang J, et al. SOAPdenovo-Trans: De novo transcriptome assembly with short RNA-Seq reads. Bioinformatics. 2014;30(12):1660-1666.
58. Piskol R, Ramaswami G, Li JB. Reliable Identification of Genomic Variants from RNA-Seq Data. Am J Hum Genet. 2013;93(4):641-651.
59. Hill JT, Demarest BL, Bisgrove BW, et al. MMAPPR: Mutation Mapping Analysis Pipeline for Pooled RNA-seq. Genome Res. 2013;23(4):687-697.
60. https://github.com/davidliwei/rnaseqmut.
61. Katz Y, Wang ET, Airoldi EM, Burge CB. Analysis and design of RNA sequencing experiments for identifying isoform regulation. Nat Methods. 2010;7(12):1009-1015.
62. Griffith M, Griffith OL, Mwenifumbo J, et al. Alternative expression analysis by RNA sequencing. Nat Methods. 2010;7(10):843-847.
63. Zhou A, Breese MR, Hao Y, et al. Alt Event Finder: a tool for extracting alternative splicing events from RNA-seq data. BMC Genomics. 2012; 13(Suppl 8):S10.
64. Anders S, Reyes A, Huber W. Detecting differential usage of exons from RNA-seq data. Genome Res. 2012;22(10):2008-2017.
65. Drewe P, Stegle O, Hartmann L, et al. Accurate detection of differential RNA processing. Nucleic Acids Res. 2013;41(10):5189-5198.
66. Mazin P, Xiong J, Liu X, et al. Widespread splicing changes in human brain development and aging. Mol Syst Biol. 2013; 9:633.
67. Rasche A, Lienhard M, Yaspo ML, Lehrach H, Herwig R. ARH-seq: identification of differential splicing in RNA-seq data. Nucl Acids Res. 2014;42(14):e110.

68. Majoros WH, Lebeck N, Ohler U, Li Song. Improved transcript isoform discovery using ORF graphs. Bioinformatics. 2014; 30(14):1958-1964.
69. Gonzàlez-Porta M, Brazma A. Identification, annotation and visualisation of extreme changes in splicing from RNA-seq experiments with SwitchSeq, bioRxiv. 2014;
70. Kim D, Salzberg SL. TopHat-Fusion: an algorithm for discovery of novel fusion transcripts. Genome Biol. 2011;12(8):R72.
71. Ge H, Liu K, Juan T, Fang F, Newman M, Hoeck W. FusionMap: detecting fusion genes from next-generation sequencing data at base-pair resolution. Bioinformatics. 2011;27(14):1922-1928.
72. Francis RW, Thompson-Wicking K, Carter KW, Anderson D, Kees UR. FusionFinder: A Software Tool to Identify Expressed Gene Fusion Candidates from RNA-Seq Data. PLoS ONE. 2012;7(6):e39987.
73. Wu H, Wu MC, Zhi D, Santorico SA, Cui X. Statistics for next generation sequencing - meeting report. Front Genet. 2012;3:128.
74. Sboner A, Habegger L, Pflueger D, Terry S, Chen DZ. FusionSeq: a modular framework for finding gene fusions by analyzing paired-end RNA-sequencing data. Genome Biol. 2010;11(10):R104.
75. Li Y, Chien J, Smith DI, Ma J. FusionHunter: identifying fusion transcripts in cancer using paired-end RNA-seq. Bioinformatics. 2011; 27(12):1708-1710.
76. Iyer MK, Chinnaiyan AM, Maher CA. ChimeraScan: a tool for identifying chimeric transcription in sequencing data. Bioinformatics. 2011;27(20):2903-2904.
77. Asmann YW, Hossain A, Necela BM, et al. A novel bioinformatics pipeline for identification and characterization of fusion transcripts in breast cancer and normal cell lines. Nucleic Acids Res. 2011; 39(15):e100.
78. McPherson A, Hormozdiari F, Zayed A, et al. deFuse: An Algorithm for Gene Fusion Discovery in Tumor RNA-Seq Data. PLoS Comput Biol. 2011;7(5):e1001138.
79. Kinsella M, Harismendy O, Nakano M, Frazer KA, Bafna V. Sensitive Gene Fusion Detection Using Ambiguously Mapping RNA-Seq Read Pairs. Bioinformatics. 2011;27(8):1068-1075.
80. Benelli M, Pescucci C, Marseglia G, Severgnini M, Torricelli F, Magi A. Discovering chimeric transcripts in paired-end RNA-seq data by using EricScript. Bioinformatics. 2012;28(24):3232-3239.
81. Wu J, Zhang W, Huang S, et al. SOAPfusion: a robust and effective computational fusion discovery tool for RNA-seq reads. Bioinformatics. 2013;29(23):2971-2978.
82. Chepelev I1, Wei G, Tang Q, Zhao K. Detection of single nucleotide variations in expressed exons of the human genome using RNA-Seq. Nucleic Acids Res. 2009;37(16):e106.
83. Atak ZK, Gianfelici V, Hulselmans G, et al. Comprehensive analysis of transcriptome variation uncovers known and novel driver events in T-cell acute lymphoblastic leukemia. PLoS Genet. 2013; 9(12):e1003997.
84. Paik S, Shak S, Tang G, et al. A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer. N Engl J Med. 2004;351(27):2817-26.
85. Sinicropi D, Qu K, Collin F, et al. Whole transcriptome RNA-Seq analysis of breast cancer recurrence risk using formalin-fixed paraffin-embedded tumor tissue. PLoS One. 2012;7(7):e40092.
86. Ma Y, Ambannavar R, Stephans J, et al. Fusion transcript discovery in formalin-fixed paraffin-embedded human breast cancer tissues reveals a link to tumor progression. PLoS One. 2014;11:9(4):e94202.
87. Cancer Genome Atlas Research Network. Genomic and epigenomic landscapes of adult de novo acute myeloid leukemia. N Engl J Med. 2013;368(22):2059-2074.
88. Rehm HL1, Bale SJ, Bayrak-Toydemir P, et al. ACMG clinical laboratory standards for next-generation sequencing. Genet Med. 2013; 15(9):733-747.
89. Gargis AS, Kalman L, Berry MW, et al. Assuring the quality of next-generation sequencing in clinical laboratory practice. Nat Biotechnol. 2012;30(11):1033-1036.
90. Ferreira-Gonzalez A, Emmadi R, Day SP, et al, Revisiting oversight and regulation of molecular-based laboratory-developed tests: a position statement of the Association for Molecular Pathology. J Mol Diagn. 2014;16(1):3-6.
91. Wu AR Neff NF, Kalisky T, et al, Quantitative assessment of single-cell RNA-sequencing methods. Nat Methods. 2014; 11(1):41-46.
92. Jaitin DA1, Kenigsberg E, Keren-Shaul H, et al. Massively parallel single-cell RNA-seq for marker-free decomposition of tissues into cell types. Science. 2014; 343(6172):776-779.
93. Lu J, Getz G, Miska EA, et al, MicroRNA expression profiles classify human cancers. Nature. 2005; 435(7043):834-838.
94. Adams BD, Kasinski AL, Slack FJ. Aberrant Regulation and Function of MicroRNAs in Cancer. Curr Biol. 2014; 24(16):R762-R776.
95. Ulitsky I, Bartel DP. lincRNAs: genomics, evolution, and mechanisms. Cell. 2013;154(1):26-46.