Invited Mini Review

From genome sequencing to the discovery of potential biomarkers in liver disease

Sumin Oh[1,2], Yeeun Jo[1], Sungju Jung[1], Sumin Yoon[1] & Kyung Hyun Yoo[1,2,]*

[1]Laboratory of Biomedical Genomics, Department of Biological Sciences, Sookmyung Women's University, Seoul 04310, [2]Research Institute of Women's Health, Sookmyung Women's University, Seoul 04310, Korea

Chronic liver disease progresses through several stages, fatty liver, steatohepatitis, cirrhosis, and eventually, it leads to hepatocellular carcinoma (HCC) over a long period of time. Since a large proportion of patients with HCC are accompanied by cirrhosis, it is considered to be an important factor in the diagnosis of liver cancer. This is because cirrhosis leads to an irreversible harmful effect, but the early stages of chronic liver disease could be reversed to a healthy state. Therefore, the discovery of biomarkers that could identify the early stages of chronic liver disease is important to prevent serious liver damage. Biomarker discovery at liver cancer and cirrhosis has enhanced the development of sequencing technology. Next generation sequencing (NGS) is one of the representative technical innovations in the biological field in the recent decades and it is the most important thing to design for research on what type of sequencing methods are suitable and how to handle the analysis steps for data integration. In this review, we comprehensively summarized NGS techniques for identifying genome, transcriptome, DNA methylome and 3D/4D chromatin structure, and introduced framework of processing data set and integrating multi-omics data for uncovering biomarkers. [BMB Reports 2020; 53(6): 299-310]

INTRODUCTION

Studies have been conducted for a long time so as to discover biomarkers at the molecular level for the diagnosis of diseases (1-3). With the development of sequencing technology, it became possible to understand the entire genes beyond the individual gene associated with a disease. High-throughput sequencing produces omics-data set and allows the identification of modifications of genome, transcriptome and epigenome.

The investigation of entire genomic DNA sequences provides individual variants called single-nucleotide polymorphisms (SNPs) and was applied so as to predict diagnosis and prognosis of a disease through analysis of genetic diversity and population genomics (4-8). The comparison of transcriptomes in various conditions can uncover disease-specific or stage-specific genes, which can be used as biomarkers. In addition, the identification of epigenomic factors and not DNA mutations, which are factors that regulates gene expression, has potential as a new biomarker. While it was possible to find biomarkers by each sequencing approach, it has recently been possible to discover high-confident biomarkers through integrative analysis of omics data. However, in order to integrate and analyze the omics data produced for each special purpose, it is necessary to understand the characteristics of the data and examine it carefully.

Liver is the largest internal organ in the body, and it has essential roles in our body such as digesting foods, detoxifying chemical, and storing energy. Chronic liver disease and cirrhosis, damaged liver states, are a cause of global mortality and morbidity (9). Liver disease could be caused by a variety of factors, such as hepatitis virus A, B, and C infection, persistent alcoholic hepatitis, and also fat accumulation in liver. Regardless of these factors, repeated injuries provoke inflammatory damage, parenchymal cell death, and matrix decomposition leading to advance fibrosis (10). Liver disease is a multi-step disease including fatty liver, steatosis, cirrhosis, and hepatocellular carcinoma (HCC). The scar matrixes typically accumulate very slowly for approximately 5-50 years, before cirrhosis, and early stages of chronic liver disease could be reversed to a healthy state (10). But once cirrhosis occurs, it becomes to have irreversible properties. It often develops complications and even progresses to cancer (10).

Thus, in the review, we summarized the overall major researches relevant to next generation sequencing (NGS) techniques from the beginning and introduce more recent studies with integrative analysis of epigenome sequencing classified by each character of omics-data, especially in liver disease (11-15).

NGS TECHNIQUES FOR DETECTING MODIFICATIONS OF DNA, CHROMATIN STRUCTURE, AND RNA

Since genome contains genetic materials of an organism, investigating the nucleotide sequence of the genome is a great way to examine the control systems that regulate cell functions. The first DNA sequencing was produced by Sanger sequencing and developed by Frederick Sanger in 1977, which was called the chain termination method (16). From this approach, the human genome project was completed and interpreting sequences of genes has been a great help in understanding human life and diseases (17, 18). However, the function of the non-coding region was not yet precisely known, which makes up the most of human genome. The development of NGS technology that overcomes the shortcomings of Sanger sequencing has provided us with a lot of information on the features of the non-coding region.

Since the new invention of NGS technique, lots of particular sequencing methods for detecting modifications of genome, transcriptome and epigenome have been introduced. In this section, of these advanced sequencing methods, the most popular ones were summarized by categories of genome, chromatin and transcriptome-based studies (Fig. 1).

NGS techniques for detecting DNA modification
Genetic variation refers to variety in gene frequencies and mutations (Fig. 1A). The first studies using NGS techniques focused on finding significant mutations as disease triggers. Typical studies using WGS and whole exome sequencing (WES) can be analyzed for detection of genetic variations and used for target sequencing of specific regions (18).

First, there are several methods for targeted sequencing. In detail, oligonucleotide-selective sequencing (OS-Seq) was developed for capturing target genome regions with high specificity analysis of cancer genomes effectively and reproducibly (19, 20). Duplex-Seq has showed increased mutation frequency levels of the small selected regions of the nuclear genome in DNA (19, 20). Repeat sequences occupy a large portion of the eukaryotic genome. Because of their distinguishable character, they have been studied in genome evolution like genomic diversity, and their role in genome have been investigated using target sequencing. For example, molecular inversion probes short tandem repeats (MIPSTR) method specifically targets short tandem repeats (STRs), which makes it possible to detect low-frequency somatic STR variants (21). The transposon insertion sequencing (TN-Seq) is a transposon sequencing that provides information about transposon insertion sites (22). In a mutant population, the sequencing can determine gene disruptions to find some of suppressors or other mutations. Retrotransposon capture sequencing (RC-Seq) is the same mechanism to the previous one, which is applied to analyze HCC samples and identify activating oncogenic pathways (23).

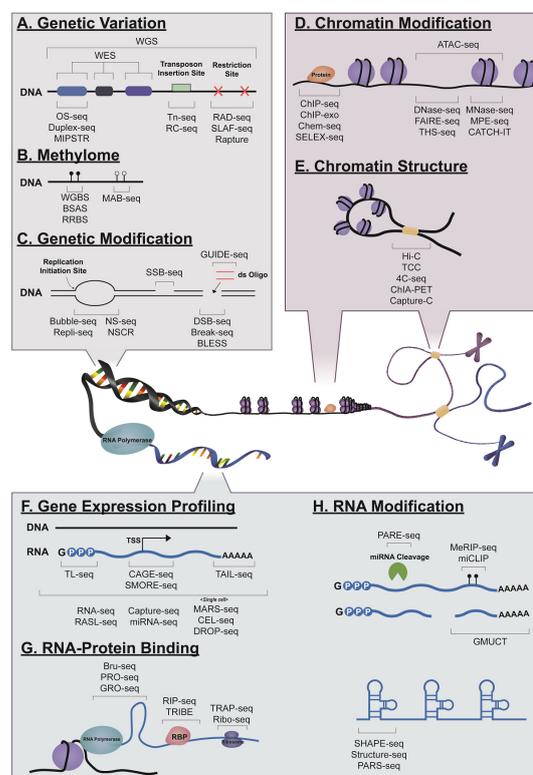SNPs and/or Single-Nucleotide Variants (SNVs) can also be

Fig. 1. Overview of next-generation sequencing techniques. (A) Detection methods for genetic variations, including in whole genome and whole exome. (B) Applications for methylation patterns. (C) Methods of genetic modifications, such as strand breaks enzyme digestion. (D) Various sequencing techniques for observing chromatin modifications. (E) Chromatin structures could be explained with sequencing applications. (F) Gene expression profiling with the form of RNAs. (G) Identification methods for RNA binding proteins and enrichment levels. (H) The formation of RNAs also detectable through sequencing applications.

detected by particular sequencing methods, which are related with restriction enzyme digesting, restriction site-associated DNA sequencing like restriction site associated DNA sequencing (RAD-Seq), specific locus amplified fragment sequencing (SLAF-Seq), and restriction site DNA capture (Rapture) (24-26).

Gene expression can also be regulated by methylation patterns on CpG regions and/or promoter regions (Fig. 1B). DNA methylation is one of the reasons of epigenetic modification, which regulates gene expression through the change of methylation and demethylation status, especially in CpG and promoter regions of the target genes. Therefore, many sequencing techniques were developed to detect the methylation pattern in genome. Whole genome bisulfite sequencing (WGBS) is the most popular tool for confirmation of methylated cytosines in whole-genomic DNA and bisulfite amplicon sequencing (BSAS). RRBS are also used to identify the methylation of DNA

(27, 28). Another method for observation of methylation in genome is the methylase assisted bisulfite sequencing (MAB-Seq), which allows quantitative mapping of both 5fC and 5caC that indicate demethylation events (29).

There are several methods for detecting genetic modification like DNA replication, and DNA strand breaks (Fig. 1C). In fundamental cellular life, DNA replication is used as important evidence for various genome regulation. Therefore, there are many techniques that have been introduced for screening the initiation sites of the DNA replication. Repli-Seq maps sequences of newly replicated DNA to the phases of cell division that validate as active DNA replication origin (30). In a similar way, Bubble-Seq, nascent strand sequencing (NS-Seq) and nascent strand capture and release (NSCR) can be utilized to verify the origin of DNA replication (31-33).

DNA strand breaks is also perceived by using sequencing techniques. The single strand break (SSB)-Sequencing shows single-strand breaks in DNA by direct detection of the pathological and physiological fracture of the DNA. On the other hand, double strand break (DSB)-Sequencing/Break-Seq/Breaks Labeling, enrichment on streptavidin and next-generation sequencing (BLESS) make it available to find double-strand breaks (DSB) on a genome wide scale (34, 35). Genome-wide unbiased identifications of DSBs evaluated sequencing (GUIDE-Seq), which is another way to detect DSB, relies on the integration of double-stranded oligodeoxynucleotides into DSBs (36).

NGS techniques for detecting dynamic chromatin structure
Physical access to DNA is an important property of chromatin that plays a crucial role in cellular characteristic (Fig. 1D). Chromatin structures can be analyzed by MNase-Seq and Methidiumpropyl-EDTA sequencing (MPE-Seq), whose techniques are based on the observation of nucleosomes and covalent attachment of tags to capture histones and identify turnover (CATCH-IT), which measures the nucleosome turnover and disruption that use metabolic labeling followed by capture of newly synthesized histones (37-39). In addition to these methods, DNase-Seq, formaldehyde assisted isolation of regulatory elements sequencing (FAIRE-Seq), and transposase hypersensitive sites sequencing (THS-Seq) could be selected to reveal genomic accessibility and open chromatin structure through representing nucleosome positioning and occupancy (40-42). Assay for transposase accessible chromatin sequencing (ATAC-Seq) relies on the hyperactive Tn5 transposase at capable regions in a genome. Proteins could bind on open chromatin regions. For this reason, active chromatin regions are used to elucidate the possibility of protein binding regions. In detail, DNA-Protein interactions could be drawn by ChIP-Seq, chromatin immunoprecipitation – exonuclease digestion (ChIP-Exo), Chem-Seq, and systematic evolution of ligands by exponential enrichment sequencing (SELEX-Seq) (43-45).

In the nucleus, the 3D structure of the genome should be related to gene expression and the importance that has been steadily increasing. Therefore, many scientists are paying attention to study and reveal chromatin looping and physical interactions (Fig. 1E). Thus, various techniques have been proceed to illustrate the structure of chromatin, such as ChIA-PET, Hi-C, Capture-C, Tethered Conformation Capture (TCC) and 4C-Seq (46-49). Chromatin interaction analysis by paired-end tag sequencing (ChIA-PET) that incorporates ChIP based technique and used for a new model of CTCF function identifying chromosome structure organization, gene transcription regulation, and linking enhancers to promoters (50).

NGS techniques for detecting RNA modification
At the level of transcription, the proceeding of measurement of gene expression status could be defined as gene expression profiling (Fig. 1F). In this step, under specific conditions, gene expression levels are usually compared to each other. For this, RNA-Seq is a commonly used technique that can examine whole transcriptome for gene expression patterns (51). RNA-mediated oligonucleotide annealing, selection and ligation sequencing (RASL-Seq) and capture-Seq are similar techniques to quantify gene expression levels (52-54). Non-coding RNAs (ncRNAs) have been unveiled as other regulators for gene expression. MicroRNAs, one of the ncRNAs, also play an essential role in the control of gene expression levels and they are detected by miRNA-Seq (55). For profiling the transcriptional diversity in a single cell, massively-parallel RNA sequencing (MARS-Seq), cell expression by linear amplification sequencing (CEL-Seq), DROP-Seq were used in the study (56, 57). Several techniques are focused on specific regions. Cap analysis gene expression sequencing (CAGE-Seq) and simultaneous mapping of RNA ends sequencing (SMORE-Seq) could be used to uncover the presence of transcription start site and measure gene expression levels (58, 59). Additionally, there are other methods for detecting specific regions of RNA. Transcript leader sequencing (TL-Seq) could be suggested for sequencing of 5′ UTR and TAIL-Seq reveals 3′ends of RNAs (60). In addition, TAIL-Seq allows to estimate Poly A tail length (61).

Some proteins have a role in RNA regulation through binding to RNA (Fig. 1G). In unwound DNA strand, RNA polymerase and some other proteins interact with it, then RNA transcripts are produced. In order to analyze that circumstance, precision nuclear run-on sequencing (PRO-Seq) presents the site of active RNA polymerase, bromouridine sequencing (Bru-Seq) and global run-on sequencing (GRO-Seq) show nascent RNA transcripts to analyze synthesis and stability of RNAs (62, 63). In addition to that, GRO-Seq has also been performed to identify enhancer RNAs (64). In order to predict the protein binding sites of RNAs, RNA immunoprecipitation sequencing (RIP-Seq), and targets of RNA binding proteins identified by editing (TRIBE) used to determine RNA-protein association and identify the target RNA sequences of RNA binding proteins (RBP) (65-67). After transcription, ribosomes interact with RNAs for protein synthesis. Ribo-Seq is a ribosome profiling technique that figures out the location of ribosome in mRNA translation (68). Translating ribosome affinity purification

sequencing (TRAP-Seq) is another method used to clarify translating mRNAs and profile cell type specific translatomes (69).

RNA methylation is another method of regulating gene expression epigenetically at the transcriptional level (Fig. 1H). Since these variations of RNA have been discovered in cancer, many studies have been carried out to affirm the methylation patterns of RNAs (70). Methylated RNA immunoprecipitation sequencing (MeRIP-Seq) was developed to show m6A methylated RNA, also miCLIP indicates m6A locations (71, 72). Degradation of RNAs can be also detected using sequencing techniques. Parallel analysis of RNA end sequencing (PARE-Seq) was published to identify microRNA cleavage sites as degrading RNA and genome wide mapping of uncapped and cleaved transcripts (GMUCT) was brought out to discover uncapped and cleaved transcripts (73, 74). Today, the secondary structure of RNA is also focalized to understand RNA modification in between processes of transcription and translation. Therefore, several techniques have been published. Selective 2'hydroxyl acylation analyzed by primer extension (SHAPE-Seq) is the RNA structure analysis technique (75). Additionally, structure-seq and parallel analysis of RNA secondary structure sequencing (PARS-Seq) are probing RNA secondary structures in genome wide scale (76, 77). These techniques can simultaneously measure single nucleotide resolution secondary and tertiary information for a lot of RNA molecules of arbitrary sequence.

DISCOVERING BIOMARKERS FOR LIVER DISEASE THROUGH NGS DATA

As liver disease could be caused by a variety of factors, such as viruses and alcohol, the treatment methods differ depending on the cause. Liver disease is developed through several stages for a long time. Unfortunately, liver disease patients are often asymptomatic and can remain unaware of their condition until late stages of the disease. Chronic liver disease is characterized by progressive hepatic fibrosis and it leads to the formation of cirrhosis, HCC, and liver failure, often requiring liver transplantation. However, it is only possible to reversibly return to a favorable state in the early stages of the disease (10). That is the reason why detection of biomarkers is needed for early diagnosis. Many studies have been conducted to find the difference between cirrhosis and cancer, which is the late stages of the disease (10). Therefore, it is necessary to discover biomarkers capable of detecting liver diseases at an early stage by comparatively analyzing specific markers for each stage of liver disease.

NGS techniques have been substantially utilized to identify functional mechanisms and novel biomarkers in diverse diseases (3, 6, 17, 78-82). In previous studies using NGS techniques, significant characters of different tissue/cell status have been identified with a single type or multiple types of NGS data in disease, development or specific condition. Biomarkers identified with NGS technique for liver disease were as summarized in Table 1. At the start, genome sequencing is one of the most

popular approaches in the identification of genomic mutations and figure out the mechanisms of diseases. Analysis of mutations using WGS or WES enables the prediction of diseases degeneration or discover influential driver genes. Marker genes have been identified by undermining essential meanings of somatic mutation patterns varied in accordance with different disease states. In progress of HCC, normal hepatocellular cells into carcinoma cells, analysis of genetic alterations using WES data was carried out to verify irregular interruption of cellular pathways related to the cancer occurrence and identify driver genes (83). Genomic variations were also observed in disease stages before tumorigenesis. In detail of the liver disease before tumorigenesis, hepatitis virus, alcohol abuse, and non-alcoholic steatohepatitis (NASH) are commonly known as the causal factors and they ultimately lead to cirrhosis, a stage of liver fibrosis (6, 84, 85). A research with WGS analysis focused on cirrhosis derived from chronic liver disease states – alcohol-related liver disease (ARLS) and non-alcoholic fatty liver disease (NAFLD) (6). They observed heterogeneity through somatic mutations and the results suggested that chronic liver disease has increased rates of mutation, complex structural variation, and low mutations targeting known HCC genes (6). In another research, the researchers were more focused on the chronic liver disease states before cirrhosis (84). Recurrent mutations on chronic liver diseases tissue were found through WES analysis (84). It provided evidence that the somatic mutations are highly related to liver fibrosis stage and specific mutations – *PKD1*, *KMT2D*, *ARID1A* and *PPARGC1B* – promote hepatic fitness and regeneration against liver injury but non-existent in cancers (84).

In addition, a large-scale study is underway to confirm the correlation between abnormality including mutation in the genome and liver disease (3, 6). For example, copy-number variations (CNVs) in 38 types of cancers were found as a part of the pan-cancer analysis of whole genomes (PCAWG) consortium analyzing 2,658 cancers and the result suggested that the CNVs could be used as diagnostic markers in the early stage of cancers (86). In addition, somatic mutations found in the hepatocellular carcinoma (HCC) were related to highly expressed hepato-specific genes, providing evidence of liver tumorigenesis (7). A research revealed that genomic markers of liver cancer could also be identified with WGS by genomic subtyping (3). In the process, they figured out the correlation between single nucleotide variations (SNVs) load and two types of heterozygosity mutations – gain-of-heterozygosity (GOH) and loss-of-heterozygosity (LOH) – by categorizing the SNV loads of 110 liver cancers obtained by paired blood-tumor WGS (3). Additionally, it was showed that the recurrent somatic survival-related CNVs (srCNVs) are linked to the LOHs, as they are more relevant to HCC short survival (3). The analysis of WGS data along with prognostic survival analysis indicated that malignant cancers tend to have a large number of SNV, LOHs, and CNV mutations (3). Based on the result, SNV load, LOH%, Signature a%, or srCNV were suggested as remark-

Table 1. Biomarkers for liver disease identified using NGS technique

Approach of NGS	Applied techniques	Biological meaning	Detected alteration	Target disease	Biomarkers	Reference
Single	WGS	Distinct relationships with environmental factors and transcription, and driver genes	Mutational and structural rearragement signatures	Liver Cancer	CTNNB1, AXIN1, PTEN, RB1, ARID2, TERT	(7)
Single	WES	Disruption of cellular pathways	Genetic alterations	HCC	TP53, CTNNB1, KEAP1, C16orf62, MLL4, RAC2	(83)
Single	WES	Recurrently mutated genes in fitness and regeneration	Accumulation of mutations	Cirrhotic livers	PKD1, KMT2D, ARID1A	(84)
Single	WGS	Early oncogenesis	Mutations in sets of driver genes	Cancer	TP53, KRAS, TERT, PIK3CA, APC, SMAD4, MEN1, DAXX, EGFR	(86)
Single	RRBS	Tumorigenesis	Hypomethylaion on Upregulated region	HCC	Mst1r, Slpi	(105)
Single	RNA-Seq	Pro-oncogenic properties such as cellular growth and proliferation, movement	CNV alterations and gene expression change	Cancer	TP53, KRAS, CTNNB1	(78)
Single	RNA-Seq	Inhibition of Growth, Migration, and Invasion of HCC	Overexpression	HCC	ZKSCAN1	(80)
Single	RNA-Seq	Treated FASN related to liver disease	Overexpression	NAFLD	PKLR, PNPLA3, PCSK9	(87)
Single	Small RNA-Seq	Controlling epithelial mesenchymal transition (EMT) and metastasis	Low expression levels	HCC and HCA	miR-200a, miR-429, miR-490-3p, miR-452, miR-766, miR-1180	(88)
Single	miRNA-Seq	Distinguishing early HCC from LC	Comparing expression by ROC curve analysis	HCC	miR-122, miR-148a	(89)
Single	miRNA-Seq	Acceleration of cell migration oand invasion	Overexpression	HBV-associated HCC	miR-21	(90)
Single	DNA-Seq	3' UTF variant related to MetS features	Genetic variations in aminotransferase loci	NAFLD	GOT2	(81)
Multiple	RNA-Seq, ChIP-Seq	Cancer cell proliferation and tumorigenesis	Demethylation of promoter is associated with the gene expression	Cancer	LIN28B	(79)
Multiple	RNA-seq, WGS	Identification of Genomic mutations and transcriptomic abberations	p53 signaling related regulation	HCC	TTK	(2)
Multiple	RNA-Seq, small RNA-Seq	Metastasis and tumorigenesis	Deregulation of lncRNAs related with DNA methylation on genomic alterations	HCC	HAND2-AS1	(91)
Multiple	Total RNA-Seq, miRNA-seq	Altered gene expression levels for multiple pathway related with cancer and non-cancer	Correlation between complex ncRNA-miRNA-mRNA network	HCC	CECR7, LINC00346, MAPKAPK5-AS1, LOC338651, FLJ90757, LOC283663	(92)
Multiple	RNA-Seq, miRNA-Seq	Induce steatosis-like phenotypes and Enhance risks of HCC	Overexpression of mRNA effects to miRNA regulatory system	Murine microsteatosis	IMP2	(93)

Table 1. Continued

Approach of NGS	Applied techniques	Biological meaning	Detected alteration	Target disease	Biomarkers	Reference
Multiple	RNA-Seq, miRNA-seq	Identification of Genomic mutations and transcriptomic abberations	Regulation of multiple metabolic pathways	HCC	TP53, AXIN1, ARID2, RPS6KA3, HNF4A, CPS1, TSC1, THRAP3	(101)
Multiple	RNA-Seq, WES	Different oncogenic pathways result in distinct tumour phenotypes	Mutation-altered gene regulation	HCC	CTNNB1, TP53, TSC1/TSC2, TERT	(102)
Multiple	RNA-Seq, DNA-Seq	Sorafenib response	Oncogene mutational burden in tumor / Overexpression	HCC	TGFa, PECAM1, NRG1	(103)
Multiple	RNA-Seq, RRBS	Observation of phenotypes in pFFC-FFC cohort	Overexpression with Hypomethylation	NASH	PDGFRβ	(104)

List of biomarkers organized by approach of NGS, application methods in liver disease. WGS: Whole Genome Sequencing; WES: Whole Exome Sequencing; RRBS: Reduced Representation Bisulfite Sequencing; RNA-Seq: RNA Sequencing; miRNA-Seq: micro RNA-Seq; ChIP-Seq: Chromatin Immunoprecipitation Sequencing; MAPS: Massive Anchored Parallel Sequencing.

able factors as genomic markers (3).

Furthermore, a transcriptome research of the liver showed differentially expressed genes between NAFLD and HCC by analyzing RNA-Seq (87). By using RNA-Seq data, co-expression analysis was performed between NAFLD and HCC by focusing FANS known as putative key regulatory gene in progressive and development to several disease stages was screened and the result confirmed that the expression levels of PCSK9, PNPLA3, and PSCK9 were associated with disease severity (87). Similarly, significant non-coding RNAs like micro RNAs for disease progression were found using sequencing techniques, such as small RNA-Seq and miRNA-Seq (88-90). These sequencing techniques are also conducted with RNA-Seq application (91-93). *IMP2* was implicated in HCC development as a risk factor through miRNA regulation (93). Deregulation of *HAND2-AS1* caused CNVs, and DMRs was revealed as a metastasis and tumorigenesis risk feature in HCC, which was confirmed with the correlation among RNA-Seq, small RNA-Seq, CNVs Affymetrix CytoscanHD array, and DNA methylation microarrays (91).

INTEGRATIVE ANALYSIS WITH TRANSCRIPTOME AND OTHER SEQUENCING APPLICATIONS

Although NGS data have been accumulated according to the advances in technical methods of sequencing, it is still not enough to uncover all of the biological phenomena. In this respect, integration of numerous NGS data types could be utilized to find further biological meanings. However, it is not a simple problem to design integrative analysis. The reason is due to the diversity of research purpose as it can even make or break the overall research. Hence, recently, how to integrate multiple sequencing data is the most remarkable point in researches discovering molecular mechanisms. Previous studies were adopted integration analysis to understand pathology of carcinoma or diseases in the liver.

In spite of previously discovered relations between mutations and diseases through genomic sequencing data analysis, there are still many limitations on understanding gene regulations relevant to diseases progression. Besides the significance of genome analysis in contribution to verify mutations, it is also essential to estimate expression levels of genes for researches on mechanisms of gene regulation. Thus, most of sequencing integration studies have been based on RNA-Seq to obtain transcriptomic information and to realize uncovered parts of different cancers or complex diseases in specific organs (82, 94-100). Through integrative analysis, previously studied theories have been confirmed with RNA-Seq adopted to other sequencing techniques, such as WGS, small RNA-Seq or miRNA-Seq, ATAC-Seq or DNase-Seq, ChIP-Seq, and/or WGBS or RRBS (Fig. 2A). The correlation between altered expression levels of genes and genetic variations with integrative analysis of RNA-Seq and Genome sequencing including WGS and WES could be confirmed (2, 13, 82, 101). As mentioned above, RNA-Seq
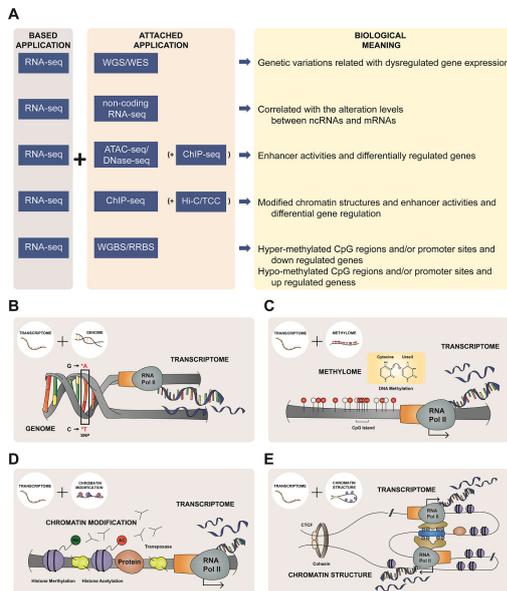
Fig. 2. Integrative analysis with complex biological meanings. (A) Integrative analysis pipeline for RNA-Seq data based and several attached optional sequencing applications with biological meanings. (B) Genome data could be integrated with transcriptome data. (C) Integrative analysis of Methylation patterns on CpG islands and RNA expression patterns could be explained with together. (D) The effect between chromatin modifications on genome and transcriptome patterns could be used to investigate gene regulation. (E) Chromatin structures involved in regulating gene expression levels.

and small RNA-Seq or miRNA-Seq that have been used to reveal the regulation of non-coding RNAs could also be associated with gene expression levels (91-93). Enhancer formations and activities are also considered with gene expression levels and this consequential meaning has been studied with RNA-Seq, ATAC-Seq or DNase-Seq, and ChIP-Seq (14). Combined of RNA-Seq, ChIP-Seq, and Hi-C or TCC data enables to explain that chromatin structural modifications and enhancer activities are components for the alterations of gene expression levels (11, 12, 98). RNA-Seq and BS related techniques, like WGBS and RRBS, can be used to find the inverse correlation between gene expression levels and methylation patterns in CpG and/or promoter regions (15, 82).

Integrative analysis of transcriptome with genome
For instance, related studies published strong correlations of genome and transcriptome (Fig. 2B) (2, 13, 101-103). These showed that somatic variations caused over-expression of oncogenes in HCC and emphasized the necessity of the integrative analysis (101). As another trial to integrate genome and transcriptome sequencing in HCC, differentially expressed genes (DEGs) were found in large CNV segments and functional analysis was performed to examine the results (2). TTK, a protein kinase related to p53 signaling, was identified with

the integrative analysis for a prognostic marker in HCC (2). Further, for therapeutic purpose, integrative analysis was performed to discover alternative drugs to sorafenib, a pre-found drug of HCC, which turned out to have limited usage due to high toxicity on HCC based on a clinical trial (13). The multi-omics analysis included genetic, transcriptomic, and additionally proteomic data of 34 liver cancer cell lines (LCCLs), including HepG2 and Huh6 of hepatoblastoma (13). Genome analysis with WES was conducted to validate similarities of genetic alterations between LCCLs and HCC and expression patterns through miRNA and mRNA analysis were integrated by elastic net regression (13). The integrated results were used to predict the sensitivity of the drugs followed by identification of molecular markers (13). In sum, integrative analysis of genetic, transcriptomic, and proteomic profiles was performed to find novel candidates of therapeutic markers in HCC (13). The result was combined with single agents, validated combinations, or drug screening, which were previously approved or being in clinical development. Thus, it provided possibilities of application of identified markers to clinical trials (13).

Integrative analysis of transcriptome with epigenome
Data integration approaches, which include epigenome but independent on DNA sequences, have been increasing in studies of various influential factors on gene regulatory mechanisms in diseases (97). Integrative analysis based on transcriptomic data was conducted with varied combinations of epigenome data for ascertaining disease progression. Environmental gene regulations could be explained with epigenetic modifications obtained by sequencing data. As a representative, the epigenetic factors include interactions of transcription factors (TFs) with specific genomic regions, DNA methylation patterns, histone acetylation or methylation, and formations of chromatin looping. We summarized previously published studies working with integrative analysis of epigenetic data based on the transcriptomic data, largely categorized as three classes – methylome, chromatin modification, and chromatin structure. In those studies, individual sequencing methods were selected depending on the particular purpose of each research.

Methylome: DNA methylation has been expected to play a central role for epigenetic changes and reported to have inverse relation with gene expression levels (Fig. 2C) (15, 82, 104, 105). There were tries to manifest the relation between the pattern of methylated genome regions and gene regulation, independent on the observed somatic mutations in HCC. A study clarified the inverse correlation between methylome landscape of C3H mice and gene expression levels through the analysis of RRBS and microarray expression analysis (105). For deeper understanding of human liver in epigenetics, sequencing data, including RNA-Seq, RRBS, DNase1-Seq, and ChIP-Seq, were analyzed (15). As an attempt to identify biomarkers in liver disease, the integration of sequencing data of epigenetic modification and transcriptome contribute to better understanding of regional tissue organization programs changing in

disease progress. The distinguishable epigenetic changes were observed based on the data among zonal networks of pericentral, intermediate, and periportal areas in different disease states of hepatocytes (15). In the point of integrated analysis of transcriptome and methylome, DEGs and differentially-methylated-regions (DMRs) of CpGs on transcriptional start sites (TSSs) were recognized using principal component analysis (PCA) (15). Gene ontology (GO) was analyzed with the DEGs representing negatively correlated expression patterns with DMRs (15). The result confirmed the certainty of functional driver genes in point zonated metabolic enzymes, which showed the consistency of hypomethylation and overexpression patterns (15). Another integrative analysis was performed to focus on an epigenetically regulated protein complex including E3 ubiquitin-like containing PHD and RING finger domain 1 (UHRF1) involved in DNA methylation and regulation on promoter regions in human hepatoblastoma (HB) cell lines – HUH6, HepT1, and HepG2 (106). Through integrative analysis of bisulfite-treated DNA pyrosequencing and RNA-Seq, epigenetic functions of UHRF1 were validated by finding HB-specific transcriptional changes on *HHIP*, *IGFBP3*, and *SFRP1* which were highly expressed with decreased methylation levels on *UHRF1* depleted HB model (106). In addition, the overexpression of *UHRF1* was shown in patients with poor disease status (106). As a result, the role of UHRF1 was identified as a critical epigenetic gene in HB and UHRF1 was suggested as a prognostic biomarker in HB (106).

Recently, integrative analysis of transcriptome and methylome was carried out to identify biomarkers in osteoporosis (82). Multi-omics data, including transcriptome, methylome, and metabolome, was integrated by sparse multiple discriminative canonical correlation analysis (SMDCCA), a multivariate integrating method that is used for searching optimal linear combination of features, and the integrated data was combined with genome (82). SMDCCA is a valid method of finding potential biomarkers. The integrated result was evaluated with pre-integrated data of 1,5994 DEGs with 1,219 DMRs and 204 DMPs to investigate potential causal effects of the pre-found biomarkers (82).

Chromatin modification: It is common to perform ChIP-Seq and combine with gene expression patterns. ChIP-Seq is the most popular method to reach out for the aim of realizing the chromatin structure in whole genome including non-coding regulatory regions (Fig. 2D). Using the technique, chromatin modifications could be observed as active and/or repressive regulatory regions through detecting histone modifications and protein bindings, such as TFs binding. Gene regulatory systems are operated on proximal and also distal regulatory regions followed by RNA transcription process. Additionally, the regulatory regions are occupied by bindings of TFs along with accessible chromatin structure. In a research of enhancers related to the Kupffer Cell (KC), DEGs of repopulating liver macrophages (RLMs) were selected by RNA-Seq analysis and the expressions of pre-defined identity genes were obtained

simultaneously (14). Open chromatin regions of the DEGs were found by ATAC-Seq and putative KC specific enhancers were identified in recruited monocytes using histone3 lysine27 acetylation (H3K27ac) ChIP-Seq, which is known as active histone marker (14). In addition to the integrative analysis of ATAC-Seq and ChIP-Seq, motif search in previously found distal open chromatin regions of RLMs was conducted and LXR, SMAD4, and RBPJ ChIP-Seq – TF ChIP-Seq – were performed (14). The result showed that liver environmental signals induce expression of regulatory related TFs, which have an influence on additional enhancers of KC differentiation (14). With more concentration on the integrative analysis of RNA-Seq and ChIP-Seq data, a recent study of gene regulations related to hepatic lipid handling compared the transcriptome and cistrome of B cell lymphoma 6 (BCL6) (98). BCL ChIP-Seq peaks were annotated to TSS and they were grouped according to their transcriptional change as 'Repressed', 'Activated', and 'Unchanged' (98). Grouped genes were analyzed by gene ontology analysis and enrichment analysis and the similar methods was selected for the PPARδ, BCL6-PPARα analysis (98). As a result, it led to conclusion that BCL6 in hepatocytes is related with repressed fatty acid oxidation and functions as a negative regulator (98).

Chromatin structure: In respect to chromatin looping structure, mapping 3D chromatin organization is a prior aim in integrative analysis of the NGS technique. As ChIP-Seq detects either cis-acting elements or trans-acting factors, the integrative analysis often involves ChIP-Seq with RNA-Seq data in addition to chromatin conformation capture sequencing data (Fig. 2E). CCCTC binding factor (CTCF) was revealed as a highly conserved zinc finger TF and it plays a key role in maintenance of topologically associating domains (TADs) relevant to establishment of chromatin structure (11). Integration of multi-omics data was processed by comparing each data to describe how CTCF contributes to the stability of TADs (11). CTCF binding regions were detected by analyzing genome-wide ChIP-Seq and gene expressions affected by the CTCF binding was obtained by integrating ChIP-Seq and RNA-Seq analyses (11). The integrated result was compared to the TAD boundaries identified by Hi-C, an NGS technique to capture the conformations of chromosome in mouse liver tissue (11). Consequently, clustered CTCF sites were found to stabilize cohesin and transcriptional regulation (11). Similarly, another previous study in mouse hepatocytes figured out the role of CTCF and cohesin and measured the effect on chromosome organization through tethered chromatin capture (TCC) (12). TCC is a modified Hi-C technique that was developed to improve mapping of low-frequency interactions on the reaction phase (49). In addition to the result, RAD21 and SMC3 ChIP-Seq were used to observe non-dividing hepatocytes when cohesin was lost by the depletion of cohesin loading factor, NIPBL/SCC2 (12). The depletion of NIPBL caused elimination of TADs even when the pattern of CTCF occupancy on chromatin boundary was not changed (12). The result suggested that the CTCF has

a distinct function with cohesin on chromosome looping (12). Moreover, the loss of NIPBL involved in dysregulation of genes meant communication impairment for gene expression between H3K27ac and H3K4me3 enriched in promoters and enhancers (12). In summary, the integrative analysis of transcriptome and chromatin conformation discovered an ability of cohesin to alter gene expressions regulated with enhancers through modification of chromatin folding (12). ChIA-PET adopted the advantage of ChIP-Seq and chromosome conformation capture system, and it could be an alternative technique to show the interactions of regulatory regions. In a previous study, CTCF-mediated ChIA-PET was used to evaluate the results from integrative analysis of Hi-C and ChIP-Seq, which was conducted to identify a super-enhancer (SE) associated with chromatin interactions, including CTCF and cohesin in K562 cell line (99). In human liver, to figure out chromatin interaction maps of gene regulatory network, disease-relevant genes involved in regulatory system were characterized by various omics data including RNA-seq, H3K4me3 and H3K27ac ChIP-seq, genome-wide associated study (GWAS), and Capture-C, all extracted from human liver tissues except Capture-C from HepG2 cell line (107). The integrative analysis of chromatin looping structure and transcriptome profiling showed the effect of enhancer activity on gene regulation like KPNB1 and SORT1 which had promoters interacted with H3K27ac peaks (107). It could be suggested that the genotype-dependent regulatory elements and driver genes related to complex trait pathogenesis could be discovered by integrative analysis of genotypes, expression levels, regulatory loci, and chromatin looping (107).

CONCLUSION

The development of NGS sequencing technique has been dramatically increased during the past two decades. Therefore, lots of different attempts have also been increased to contrive the improved techniques to specific purposes. As a result, various kinds of methods using NGS have been come into the world to discover the whole process of genomic and epigenetic regulation on biological phenomena in more details. Based on the discoveries, identification of novel biomarkers gets another approach. In contrast to previous biomarkers uncovered by radiology, sequencing data provide more convinced evidences in a microscopic aspect. Although trends on novel biomarkers of diverse diseases increase with abundance of biological data from sequencing, the procedure of data analysis has not yet been constructed systemically. Even with same type of data, integration depends considerably on analysis tools (100). Therefore, integration analysis is still figuring it out with careful consideration of how to integrate various sequencing data that have different properties and how to deal with huge size of sequencing data. In most of the previous studies, traditional statistical methods have conducted research on data analysis. However, following the generation speed and trend

of bulk size sequencing data, nowadays, many studies have aggressively implemented to apply computational algorithms, such as AI (97). AI is an upcoming big trend in data analysis field especially with classification. With the results of analysis data using NGS as an input, iterative modeling process of AI makes it possible to classify samples into several disease stages and it can also suggest significant genes as biomarkers. The advanced integrative analysis of NGS and the modeling more elaborate AI algorithms will let us discover novel biomarkers unseen before.

ACKNOWLEDGEMENTS

CONFLICTS OF INTEREST

The authors have no conflicting interests.

REFERENCES

1. Luo P, Yin P, Hua R et al (2018) A Large-scale, multi-center serum metabolite biomarker identification study for the early detection of hepatocellular carcinoma. Hepatology 67, 662-675
2. Miao R, Luo H, Zhou H et al (2014) Identification of prognostic biomarkers in hepatitis B virus-related hepatocellular carcinoma and stratification by integrative multi-omics analysis. J Hepatol 61, 840-849
3. Wu Z, Long X, Tsang SY et al (2020) Genomic subtyping of liver cancers with prognostic application. BMC Cancer 20, 84
4. Tahmasebi S, Khoutorsky A, Mathews MB and Sonenberg N (2018) Translation deregulation in human disease. Nat Rev Mol Cell Biol 19, 791-807
5. Ng SB, Turner EH, Robertson PD et al (2009) Targeted capture and massively parallel sequencing of 12 human exomes. Nature 461, 272-276
6. Brunner SF, Roberts ND, Wylie LA et al (2019) Somatic mutations and clonal dynamics in healthy and cirrhotic human liver. Nature 574, 538-542
7. Letouze E, Shinde J, Renault V et al (2017) Mutational signatures reveal the dynamic interplay of risk factors and cellular processes during liver tumorigenesis. Nat Commun 8, 1315
8. Genomes Project C, Abecasis GR, Altshuler D et al (2010) A map of human genome variation from population-scale sequencing. Nature 467, 1061-1073
9. Sepanlou SG, Safiri S, Bisignano C et al (2020) The global, regional, and national burden of cirrhosis by cause in 195 countries and territories, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017. Lancet Gastroenterol Hepatol 5, 245-266
10. Pellicoro A, Ramachandran P, Iredale JP and Fallowfield

JA (2014) Liver fibrosis and repair: immune regulation of wound healing in a solid organ. Nat Rev Immunol 14, 181-194

11. Kentepozidou E, Aitken SJ, Feig C et al (2020) Clustered CTCF binding is an evolutionary mechanism to maintain topologically associating domains. Genome Biol 21, 5

12. Schwarzer W, Abdennur N, Goloborodko A et al (2017) Two independent modes of chromatin organization revealed by cohesin removal. Nature 551, 51-56

13. Caruso S, Calatayud AL, Pilet J et al (2019) Analysis of Liver Cancer Cell Lines Identifies Agents With Likely Efficacy Against Hepatocellular Carcinoma and Markers of Response. Gastroenterology 157, 760-776

14. Sakai M, Troutman TD, Seidman JS et al (2019) Liver-Derived Signals Sequentially Reprogram Myeloid Enhancers to Initiate and Maintain Kupffer Cell Identity. Immunity 51, 655-670 e658

15. Brosch M, Kattler K, Herrmann A et al (2018) Epigenomic map of human liver reveals principles of zonated morphogenic and metabolic control. Nat Commun 9, 4150

16. Sanger F, Nicklen S and Coulson AR (1977) DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci U S A 74, 5463-5467

17. Turnbull C, Scott RH, Thomas E et al (2018) The 100 000 Genomes Project: bringing whole genome sequencing to the NHS. BMJ 361, k1687

18. Berger MF and Mardis ER (2018) The emerging clinical relevance of genomics in cancer medicine. Nat Rev Clin Oncol 15, 353-365

19. Myllykangas S, Buenrostro JD, Natsoulis G, Bell JM and Ji HP (2011) Efficient targeted resequencing of human germline and cancer genomes by oligonucleotide-selective sequencing. Nat Biotechnol 29, 1024-1027

20. Ahn EH and Lee SH (2019) Detection of Low-Frequency Mutations and Identification of Heat-Induced Artifactual Mutations Using Duplex Sequencing. Int J Mol Sci 20

21. Carlson KD, Sudmant PH, Press MO, Eichler EE, Shendure J and Queitsch C (2015) MIPSTR: a method for multiplex genotyping of germline and somatic STR variation across many individuals. Genome Res 25, 750-761

22. van Opijnen T, Bodi KL and Camilli A (2009) Tn-seq: high-throughput parallel sequencing for fitness and genetic interaction studies in microorganisms. Nat Methods 6, 767-772

23. Schauer SN, Carreira PE, Shukla R et al (2018) L1 retrotransposition is a common feature of mammalian hepatocarcinogenesis. Genome Res 28, 639-653

24. Davey JW and Blaxter ML (2010) RADSeq: next-generation population genetics. Brief Funct Genomics 9, 416-423

25. Sun X, Liu D, Zhang X et al (2013) SLAF-seq: an efficient method of large-scale de novo SNP discovery and genotyping using high-throughput sequencing. PLoS One 8, e58700

26. Ali OA, O'Rourke SM, Amish SJ et al (2016) RAD Capture (Rapture): Flexible and Efficient Sequence-Based Genotyping. Genetics 202, 389-400

27. Masser DR, Berg AS and Freeman WM (2013) Focused, high accuracy 5-methylcytosine quantitation with base resolution by benchtop next-generation sequencing. Epigenetics Chromatin 6, 33

28. Meissner A, Gnirke A, Bell GW, Ramsahoye B, Lander ES and Jaenisch R (2005) Reduced representation bisulfite sequencing for comparative high-resolution DNA methylation analysis. Nucleic Acids Res 33, 5868-5877

29. Neri F, Incarnato D, Krepelova A, Parlato C and Oliviero S (2016) Methylation-assisted bisulfite sequencing to simultaneously map 5fC and 5caC on a genome-wide scale for DNA demethylation analysis. Nat Protoc 11, 1191-1205

30. Hansen RS, Thomas S, Sandstrom R et al (2010) Sequencing newly replicated DNA reveals widespread plasticity in human replication timing. Proc Natl Acad Sci U S A 107, 139-144

31. Mesner LD, Valsakumar V, Cieslik M, Pickin R, Hamlin JL and Bekiranov S (2013) Bubble-seq analysis of the human genome reveals distinct chromatin-mediated mechanisms for regulating early- and late-firing origins. Genome Res 23, 1774-1788

32. Foulk MS, Urban JM, Casella C and Gerbi SA (2015) Characterizing and controlling intrinsic biases of lambda exonuclease in nascent strand sequencing reveals phasing between nucleosomes and G-quadruplex motifs around a subset of human replication origins. Genome Res 25, 725-735

33. Kunnev D, Freeland A, Qin M, Wang J and Pruitt SC (2015) Isolation and sequencing of active origins of DNA replication by nascent strand capture and release (NSCR). J Biol Methods 2, e33

34. Baranello L, Kouzine F, Wojtowicz D et al (2014) DNA break mapping reveals topoisomerase II activity genome-wide. Int J Mol Sci 15, 13111-13122

35. Crosetto N, Mitra A, Silva MJ et al (2013) Nucleotide-resolution DNA double-strand break mapping by next-generation sequencing. Nat Methods 10, 361-365

36. Tsai SQ, Zheng Z, Nguyen NT et al (2015) GUIDE-seq enables genome-wide profiling of off-target cleavage by CRISPR-Cas nucleases. Nat Biotechnol 33, 187-197

37. Schones DE, Cui K, Cuddapah S et al (2008) Dynamic regulation of nucleosome positioning in the human genome. Cell 132, 887-898

38. Ishii H, Kadonaga JT and Ren B (2015) MPE-seq, a new method for the genome-wide analysis of chromatin structure. Proc Natl Acad Sci U S A 112, E3457-3465

39. Deal RB, Henikoff JG and Henikoff S (2010) Genome-wide kinetics of nucleosome turnover determined by metabolic labeling of histones. Science 328, 1161-1164

40. Giresi PG, Kim J, McDaniell RM, Iyer VR and Lieb JD (2007) FAIRE (Formaldehyde-Assisted Isolation of Regulatory Elements) isolates active regulatory elements from human chromatin. Genome Res 17, 877-885

41. Sos BC, Fung HL, Gao DR et al (2016) Characterization of chromatin accessibility with a transposome hypersensitive sites sequencing (THS-seq) assay. Genome Biol 17, 20

42. Klein DC and Hainer SJ (2019) Genomic methods in profiling DNA accessibility and factor localization. Chromosome Res 28, 69-85

43. Zentner GE and Henikoff S (2014) High-resolution digital

profiling of the epigenome. Nat Rev Genet 15, 814-827

44. Anders L, Guenther MG, Qi J et al (2014) Genome-wide localization of small molecules. Nat Biotechnol 32, 92-96

45. Riley TR, Slattery M, Abe N et al (2014) SELEX-seq: a method for characterizing the complete repertoire of binding site preferences for transcription factor complexes. Methods Mol Biol 1196, 255-278

46. Li G, Cai L, Chang H et al (2014) Chromatin Interaction Analysis with Paired-End Tag (ChIA-PET) sequencing technology and application. BMC Genomics 15 Suppl 12, S11-S11

47. Harewood L, Kishore K, Eldridge MD et al (2017) Hi-C as a tool for precise detection and characterisation of chromosomal rearrangements and copy number variation in human tumours. Genome Biol 18, 125

48. Davies JO, Telenius JM, McGowan SJ et al (2016) Multiplexed analysis of chromosome conformation at vastly improved sensitivity. Nat Methods 13, 74-80

49. Kalhor R, Tjong H, Jayathilaka N, Alber F and Chen L (2011) Genome architectures revealed by tethered chromosome conformation capture and population-based modeling. Nat Biotechnol 30, 90-98

50. Li G, Ruan X, Auerbach RK et al (2012) Extensive promoter-centered chromatin interactions provide a topological basis for transcription regulation. Cell 148, 84-98

51. Wang Z, Gerstein M and Snyder M (2009) RNA-Seq: a revolutionary tool for transcriptomics. Nat Rev Genet 10, 57-63

52. Li H, Qiu J and Fu XD (2012) RASL-seq for massively parallel and quantitative analysis of gene expression. Curr Protoc Mol Biol Chapter 4, Unit 4 13 11-19

53. Mercer TR, Clark MB, Crawford J et al (2014) Targeted sequencing for gene discovery and quantification using RNA CaptureSeq. Nat Protoc 9, 989-1009

54. Routh A, Head SR, Ordoukhanian P and Johnson JE (2015) ClickSeq: Fragmentation-Free Next-Generation Sequencing via Click Ligation of Adaptors to Stochastically Terminated 3'-Azido cDNAs. J Mol Biol 427, 2610-2616

55. Motameny S, Wolters S, Nurnberg P and Schumacher B (2010) Next Generation Sequencing of miRNAs - Strategies, Resources and Methods. Genes (Basel) 1, 70-84

56. Hashimshony T, Wagner F, Sher N and Yanai I (2012) CEL-Seq: single-cell RNA-Seq by multiplexed linear amplification. Cell Rep 2, 666-673

57. Macosko EZ, Basu A, Satija R et al (2015) Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets. Cell 161, 1202-1214

58. Kawaji H, Lizio M, Itoh M et al (2014) Comparison of CAGE and RNA-seq transcriptome profiling using clonally amplified and single-molecule next-generation sequencing. Genome Res 24, 708-717

59. Park D, Morris AR, Battenhouse A and Iyer VR (2014) Simultaneous mapping of transcript ends at single-nucleotide resolution and identification of widespread promoter-associated non-coding RNA governed by TATA elements. Nucleic Acids Res 42, 3736-3749

60. Arribere JA and Gilbert WV (2013) Roles for transcript leaders in translation and mRNA decay revealed by transcript leader sequencing. Genome Res 23, 977-987

61. Chang H, Lim J, Ha M and Kim VN (2014) TAIL-seq: genome-wide determination of poly(A) tail length and 3' end modifications. Mol Cell 53, 1044-1052

62. Mahat DB, Kwak H, Booth GT et al (2016) Base-pair-resolution genome-wide mapping of active RNA polymerases using precision nuclear run-on (PRO-seq). Nat Protoc 11, 1455-1476

63. Paulsen MT, Veloso A, Prasad J et al (2013) Coordinated regulation of synthesis and stability of RNA during the acute TNF-induced proinflammatory response. Proc Natl Acad Sci U S A 110, 2240-2245

64. Franco HL, Nagari A, Malladi VS et al (2018) Enhancer transcription reveals subtype-specific gene expression programs controlling breast cancer pathogenesis. Genome Res 28, 159-170

65. Chu C, Qu K, Zhong FL, Artandi SE and Chang HY (2011) Genomic maps of long noncoding RNA occupancy reveal principles of RNA-chromatin interactions. Mol Cell 44, 667-678

66. Zhao J, Ohsumi TK, Kung JT et al (2010) Genome-wide identification of polycomb-associated RNAs by RIP-seq. Mol Cell 40, 939-953

67. McMahon AC, Rahman R, Jin H et al (2016) TRIBE: Hijacking an RNA-Editing Enzyme to Identify Cell-Specific Targets of RNA-Binding Proteins. Cell 165, 742-753

68. Ingolia NT, Ghaemmaghami S, Newman JRS and Weissman JS (2009) Genome-Wide Analysis in Vivo of Translation with Nucleotide Resolution Using Ribosome Profiling. Science 324, 218

69. Jiao Y and Meyerowitz EM (2010) Cell-type specific analysis of translating RNAs in developing flowers reveals new levels of control. Mol Syst Biol 6, 419

70. Delaunay S and Frye M (2019) RNA modifications regulating cell fate in cancer. Nat Cell Biol 21, 552-559

71. Panneerdoss S, Eedunuri VK, Yadav P et al (2018) Cross-talk among writers, readers, and erasers of m(6)A regulates cancer growth and progression. Science Advances 4, eaar8263-eaar8263

72. Linder B, Grozhik AV, Olarerin-George AO, Meydan C, Mason CE and Jaffrey SR (2015) Single-nucleotide-resolution mapping of m6A and m6Am throughout the transcriptome. Nat Methods 12, 767-772

73. German MA, Luo S, Schroth G, Meyers BC and Green PJ (2009) Construction of Parallel Analysis of RNA Ends (PARE) libraries for the study of cleaved miRNA targets and the RNA degradome. Nat Protoc 4, 356-362

74. Yu X, Willmann MR, Anderson SJ and Gregory BD (2016) Genome-Wide Mapping of Uncapped and Cleaved Transcripts Reveals a Role for the Nuclear mRNA Cap-Binding Complex in Cotranslational RNA Decay in Arabidopsis. Plant Cell 28, 2385-2397

75. Lucks JB, Mortimer SA, Trapnell C et al (2011) Multiplexed RNA structure characterization with selective 2'-hydroxyl acylation analyzed by primer extension sequencing (SHAPE-Seq). Proc Natl Acad Sci U S A 108, 11063-11068

76. Fang R, Moss WN, Rutenberg-Schoenberg M and Simon MD (2015) Probing Xist RNA Structure in Cells Using Targeted Structure-Seq. PLoS Genet 11, e1005668

77. Wan Y, Qu K, Ouyang Z and Chang HY (2013) Genome-wide mapping of RNA structure using nuclease digestion and high-throughput sequencing. Nat Protoc 8, 849-869

78. Castven D, Becker D, Czauderna C et al (2019) Application of patient-derived liver cancer cells for phenotypic characterization and therapeutic target identification. Int J Cancer 144, 2782-2794

79. Guo W, Hu Z, Bao Y et al (2018) A LIN28B Tumor-Specific Transcript in Cancer. Cell Rep 22, 2016-2025

80. Yao Z, Luo J, Hu K et al (2017) ZKSCAN1 gene and its related circular RNA (circZKSCAN1) both inhibit hepatocellular carcinoma cell growth, migration, and invasion but through different signaling pathways. Mol Oncol 11, 422-437

81. Sookoian S, Castano GO, Scian R et al (2016) Serum aminotransferases in nonalcoholic fatty liver disease are a signature of liver metabolic perturbations at the amino acid and Krebs cycle level. Am J Clin Nutr 103, 422-434

82. Qiu C, Yu F, Su K et al (2020) Multi-omics Data Integration for Identifying Osteoporosis Biomarkers and Their Biological Interaction and Causal Mechanisms. iScience 23, 100847

83. Cleary SP, Jeck WR, Zhao X et al (2013) Identification of driver genes in hepatocellular carcinoma by exome sequencing. Hepatology 58, 1693-1702

84. Zhu M, Lu T, Jia Y et al (2019) Somatic Mutations Increase Hepatic Clonal Fitness and Regeneration in Chronic Liver Disease. Cell 177, 608-621 e612

85. Singal AG and El-Serag HB (2015) Hepatocellular Carcinoma From Epidemiology to Prevention: Translating Knowledge into Practice. Clin Gastroenterol Hepatol 13, 2140-2151

86. Gerstung M, Jolly C, Leshchiner I et al (2020) The evolutionary history of 2,658 cancers. Nature 578, 122-128

87. Lee S, Zhang C, Liu Z et al (2017) Network analyses identify liver-specific targets for treating liver diseases. Mol Syst Biol 13, 938

88. Zheng J, Sadot E, Vigidal JA et al (2018) Characterization of hepatocellular adenoma and carcinoma using microRNA profiling and targeted gene sequencing. PLoS One 13, e0200776

89. Wang Y, Zhang C, Zhang P et al (2018) Serum exosomal microRNAs combined with alpha-fetoprotein as diagnostic markers of hepatocellular carcinoma. Cancer Med 7, 1670-1679

90. Mizuguchi Y, Mishima T, Yokomuro S et al (2011) Sequencing and bioinformatics-based analyses of the microRNA transcriptome in hepatitis B-related hepatocellular carcinoma. PLoS One 6, e15304

91. Yang Y, Chen L, Gu J et al (2017) Recurrently deregulated lncRNAs in hepatocellular carcinoma. Nat Commun 8, 14421

92. Zhang J, Fan D, Jian Z, Chen GG and Lai PB (2015) Cancer Specific Long Noncoding RNAs Show Differential Expression Patterns and Competing Endogenous RNA Potential in Hepatocellular Carcinoma. PLoS One 10, e0141042

93. Dehghani Amirabad A, Ramasamy P, Wierz M et al (2018) Transgenic expression of the RNA binding protein IMP2 stabilizes miRNA targets in murine microsteatosis. Biochim Biophys Acta Mol Basis Dis 1864, 3099-3108

94. Lu M and Zhan X (2018) The crucial role of multiomic approach in cancer research and clinically relevant outcomes. EPMA J 9, 77-102

95. Gallo Cantafio ME, Grillone K, Caracciolo D et al (2018) From Single Level Analysis to Multi-Omics Integrative Approaches: A Powerful Strategy towards the Precision Oncology. High Throughput 7, 33

96. Schulze K, Nault JC and Villanueva A (2016) Genetic profiling of hepatocellular carcinoma using next-generation sequencing. J Hepatol 65, 1031-1042

97. Cazaly E, Saad J, Wang W, Heckman C, Ollikainen M and Tang J (2019) Making Sense of the Epigenome Using Data Integration Approaches. Front Pharmacol 10, 126

98. Sommars MA, Ramachandran K, Senagolage MD et al (2019) Dynamic repression by BCL6 controls the genome-wide liver response to fasting and steatosis. Elife 8, e43922

99. Huang J, Li K, Cai W et al (2018) Dissecting super-enhancer hierarchy based on chromatin interactions. Nat Commun 9, 943

100. Koohy H, Down TA, Spivakov M and Hubbard T (2014) A comparison of peak callers used for DNase-Seq data. PLoS One 9, e96303

101. Shiraishi Y, Fujimoto A, Furuta M et al (2014) Integrated analysis of whole genome and transcriptome sequencing reveals diverse transcriptomic aberrations driven by somatic genomic changes in liver cancers. PLoS One 9, e114263

102. Calderaro J, Couchy G, Imbeaud S et al (2017) Histological subtypes of hepatocellular carcinoma are related to gene mutations and molecular tumour classification. J Hepatol 67, 727-738

103. Sakai K, Takeda H, Nishijima N et al (2015) Targeted DNA and RNA sequencing of fine-needle biopsy FFPE specimens in patients with unresectable hepatocellular carcinoma treated with sorafenib. Oncotarget 6, 21636-21644

104. Gutierrez Sanchez LH, Tomita K, Guo Q et al (2018) Perinatal Nutritional Reprogramming of the Epigenome Promotes Subsequent Development of Nonalcoholic Steatohepatitis. Hepatol Commun 2, 1493-1512

105. Matsushita J, Okamura K, Nakabayashi K et al (2018) The DNA methylation profile of liver tumors in C3H mice and identification of differentially methylated regions involved in the regulation of tumorigenic genes. BMC Cancer 18, 317

106. Beck A, Trippel F, Wagner A et al (2018) Overexpression of UHRF1 promotes silencing of tumor suppressor genes and predicts outcome in hepatoblastoma. Clin Epigenetics 10, 27

107. Caliskan M, Manduchi E, Rao HS et al (2019) Genetic and Epigenetic Fine Mapping of Complex Trait Associated Loci in the Human Liver. Am J Hum Genet 105, 89-107