BioTechniques
The International Journal of Life Science Methods

Lab essentials:
Next-generation sequencing

Taylor & Francis
www.biotechniques.com

ABOUT

This eBook explores next-generation sequencing (NGS) methods and their applications in neurological and cancer research, with key content from *BioTechniques'* digital hub and Taylor & Francis journals, *BioTechniques*, *Epigenomics* and *Personalized Medicine*. In this eBook, you'll find a DNA sequencing guide as well as a Tech News and webinar discussing the use of long-read sequencing in microbiome research.

CONTENTS

Long-read sequencing expedites rare disease diagnoses

By Beatrice Bowlby (Digital Editor)

Long-read sequencing is being harnessed to improve the rate of rare disease diagnosis.

Researchers at the University of California Santa Cruz (CA, USA) have demonstrated that long-read sequencing (LRS) has the potential to expedite rare disease diagnoses with a single test and at a lower cost. Reducing the time to diagnosis from years to days for individuals with rare diseases would mean earlier intervention and, therefore, better outcomes.

Worldwide, one in ten people is impacted by a rare genetic disease; however, 50% of those cases are undiagnosed. This is in part due to the type of testing being conducted, which utilizes short-read sequencing (SRS) to investigate the genome. SRS looks for genetic variants that cause disease, reading the genetic code in sequences of 150–250 bases at a time. The problems with SRS in this context are that it can't provide a comprehensive view of the genome due to its short reading frame, and it can't reveal from which parent a variant was inherited, a process called phasing. These limitations can lead healthcare professionals to miss crucial evidence that could help them make a diagnosis. This method may also require multiple tests, which drive up costs and extend the time to diagnosis.

In the current study, the team investigated the potential of LRS to overcome the limitations of SRS. LRS is a next-generation sequencing method that can read longer strands of DNA as well as provide phasing and methylation information. "Rare diseases are something that people have been struggling to diagnose for so many years, and if we have a sequencing technology which streamlines diagnostic testing, I think that will be a huge contribution – and that is what we tested as part of this paper," commented first author Shloka Negi.

The study focused on rare diseases caused by disruption to a single gene. Working with 42 individuals with rare diseases, the researchers used innovative LRS – nanopore sequencing – to glean highly accurate genetic information from each person for US$1000 per sample. These reads were then analyzed against a large, complete reference genome to identify variants and collect phasing and methylation

data using the Napu pipeline, a wet lab and computational protocol developed in co-lead author Benedict Paten's lab.

They found that LRS provided a more exhaustive dataset compared to short-read data, conclusively diagnosing 11 of the 42 individuals using a single, rapid and cost-effective protocol. Of the 11 diagnoses, four were congenital adrenal hypoplasia, a rare disease that affects adrenal gland function and is caused by a gene located in a particularly difficult-to-access region of the genome, making current diagnostic tests for this genetic disease ineffective.

"To solve these cases, we developed a new pangenomic tool that integrates new high-quality assemblies like the 'telomere-to-telomere' reference genome," reported co-lead author Jean Monlong. "... We know many rare diseases involve regions of the human genome that have been historically difficult to study, so our results encourage us to extend our approach to more of those diseases that have been at a standstill for a long time."

Reference:

Negi S, Stenton SL, Berger SI *et al.* Advancing long-read nanopore genome assembly and accurate variant calling for rare disease detection. *Am. J. Hum. Genet.* doi:10.1016/j.ajhg.2025.01.002 (2025) (Epub ahead of print).

Click here to view the press release.

DNA sequencing reveals cause of rare neurological disorder

By Annie Coulson (Digital Editor)

After a 25-year-long search, a new sequencing technique has helped researchers pinpoint the genetic cause of SCA4, a rare movement disorder.

Spinocerebellar ataxia 4 (SCA4) is a rare genetic condition that severely impacts the lives of those affected by the disease. Like many other rare diseases, there is no known cure for SCA4, and until now, no known cause. Now, after 25 years of searching, a multi-institutional group of researchers led by a team at the University of Utah (UT, USA) has discovered the genetic difference that causes SCA4, providing families with much-needed answers and offering hope for potential treatments.

SCA4 is a progressive movement disorder, with symptoms that usually appear in an individual's 40s or 50s but can start as early as the late teens in some people. The pattern of inheritance of SCA4 makes it clear that the disease is genetic, and a study in 1996 situated the gene responsible to a specific region of one chromosome. However, repeated segments and an unusual chemical makeup made the region difficult to analyze, preventing researchers from pinpointing the mutation for over 25 years.

To overcome this, the Utah-led group used a recently developed technique, long-read single-strand whole-genome sequencing (LR-GS), to sequence DNA from affected and unaffected people from several families in Utah. They found that in individuals with SCA4, a section of the ZFHX3 gene is much longer than it should be, containing a long string of repetitive DNA called a GGC-repeat expansion.

To investigate this further, the researchers isolated cells from people with SCA4 and found ZFHX3–p62–ubiquitin protein aggregates. They also examined fibroblasts and induced pluripotent stem cells with the repetitive version of ZFHX3 and found increased ZFHX3 protein levels that can gum up the cells' protein-recycling machinery. "This mutation is a toxic expanded repeat and we think that it actually jams up how a cell deals with unfolded or misfolded proteins," said co-corresponding author Stefan Pulst.

A similar phenomenon occurs in another form of ataxia, SCA2. The same group is currently testing a potential therapy for SCA2 in clinical trials, and the similarities

between the two conditions raise the possibility that the treatment might benefit patients with SCA4 as well.

Establishing the genetic cause of SCA4 provides individuals with answers and, hopefully, effective treatments in the future. "The only step to really improve the life of patients with inherited disease is to find out what the primary cause is. We now can attack the effects of this mutation potentially at multiple levels," concluded Pulst.

Reference:

Figueroa KP, Gross C, Buena-Atienza E *et al.* A GGC-repeat expansion in ZFHX3 encoding polyglycine causes spinocerebellar ataxia type 4 and impairs autophagy. *Nat. Genet.* doi: 10.1038/s41588-024-01719-5 (2024) (Epub ahead of print).

Click here to view the press release.

GUIDE

Ten tips for high-throughput DNA sequencing

BY EMMA WATERS, QUADRAM INSTITUTE (NORWICH, UK)

DNA sequencing is a revolutionary technology that enables scientists to decipher the life-giving instructions contained within the genetic code of all organisms. At its core, DNA sequencing is the process of determining the precise order of nucleotide bases (adenine, cytosine, guanine and thymine) within a DNA molecule. This information provides a blueprint of the genetic instructions encoded in the DNA, allowing scientists to explore the structure, function and variability of genes and genomes.

To assist researchers in harnessing the full potential of high-throughput DNA sequencing, here's a top ten tip guide.

Introduction

DNA sequencing has transformed the field of biology. In medical microbiology, it has significantly advanced our understanding of genetic diseases, enabling the identification of disease-causing mutations and the development of personalized treatments. In evolutionary biology, it has shed light on the history and relationships between species, providing insights into the mechanisms of adaptation and speciation. Moreover, DNA sequencing has opened new frontiers in fields like forensic science, agriculture, conservation biology and microbial ecology, to name just a few.

The history of DNA sequencing is a remarkable journey that spans the last few decades. It all began in the 1970s when Frederick Sanger pioneered the first short-read DNA sequencing technique. This technique was used to sequence the first full genome of a virus, which was around 5000 bases long. In the 1980s, efforts were made to automate Sanger sequencing, which greatly increased the speed and efficiency, enabling the sequencing of larger genomes. These advancements allowed the launch of the Human Genome Project in 1990, where a group of scientists wanted to determine the 3 billion bases of the human genome, which was only achieved 13 years later! During that time, other scientists sequenced the genomes of different organisms, celebrating each time they sequenced a bigger genome; from 1995 when the first bacterial genome was sequenced, containing a couple of million bases [1], to 5 years later when a fly genome was sequenced, which contained a couple of hundred million bases [2].

Over this period, sequencing and computer power continued to improve, allowing for massive parallel sequencing of DNA fragments and thus the introduction of next-generation sequencing (NGS) technologies. These advancements – made only 7 years after the original Human Genome Project ended – facilitated the sequencing of thousands of human genomes and marked the start of understanding genetic diseases. In 2014, Public Health England started to use short-read sequencing at a national scale to

manage infectious diseases, particularly for tuberculosis and bacterial foodborne diseases, as this type of sequencing is highly accurate and can rapidly identify single nucleotide polymorphisms (SNPs) [3]. However, short-read sequencing, which sequences DNA fragments of hundreds of bases in length, is unable to resolve long-repeat sequences found in genomes that can span over several thousand bases.

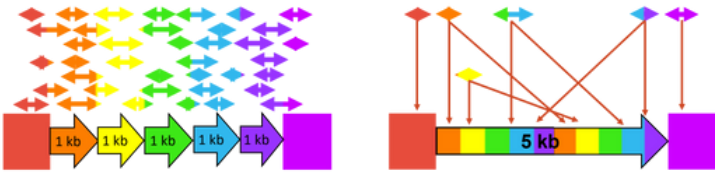


Figure 1: (Left) Short-reads of 100-1000 bases in length, mapped onto a section of genome containing easy-to-resolve genes. (Right) The mapping of short-reads is confused by sections of genomes that contain long-repeat sequences, like those of ribosomal operons found in all living organisms.

In 2009, the introduction of third-generation sequencing technologies brought about a further breakthrough in DNA sequencing [4]. These long-read sequencing techniques can sequence much longer DNA fragments (thousands to millions of bases long) and assemble genomes with increased accuracy whilst identifying large structural mutations where large genomic fragments rearrange within genomes and can shuffle the order of genetic instructions [5].

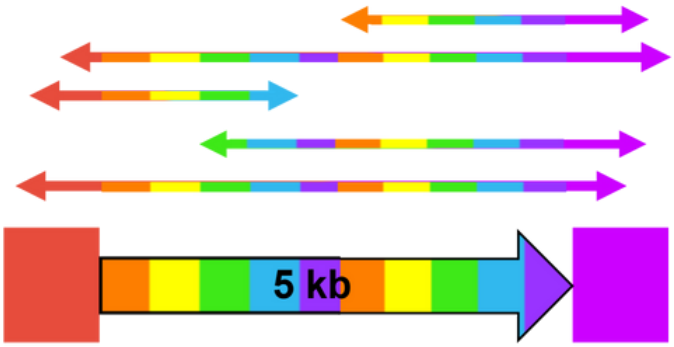
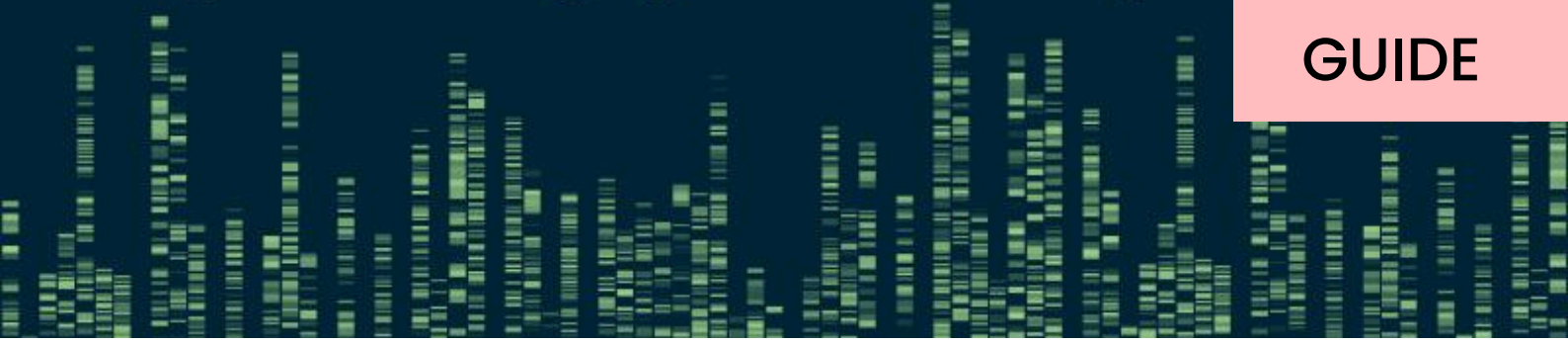


Figure 2: Long-read sequences mapped onto and resolving a section of a genome, which contains long-repeat sequences.

In this era of genomics, DNA sequencing continues to evolve rapidly. Sequencing technologies and their associated library preparation have become more accurate, faster, cost-effective and portable, leading to the potential for DNA sequencing to be performed on anything, by anyone and anywhere.

Tip 1: Choose the appropriate sequencing platform

Consider desired requirements like read length, error rate and data output to match your needs. Generally, if you are interested in SNPs and not worried about complete assemblies, short-read platforms like Illumina's MiSeq, NextSeq and NovaSeq offer high-throughput capabilities. However, if you are interested in structural variation, you will require long-read sequencing capabilities such as PacBio's Sequel II or Revio systems, or Oxford Nanopore's MinION or PromethION. You could also perform both short- and long-read sequencing to generate the gold standard, hybrid assemblies.



Tip 2: Optimize DNA extraction

Optimize sample preparation and DNA extraction for your desired sequencing method. If you are performing long-read sequencing, use appropriate high-molecular-weight DNA extraction protocols to minimize DNA shearing and degradation, as fragmented DNA can impact long-read sequencing performance. Quality control measures, such as checking DNA integrity via gel electrophoresis or fluorometric quantification, should be implemented to ensure the samples meet the sequencing requirements.

Tip 3: Choose the appropriate library preparation

Follow optimized library preparation protocols specific to your chosen sequencing platform. These protocols generally follow the same outlines: DNA fragmentation and size selection (which might be omitted in long-read methods), end repair, barcode ligation for multiplexing, adapter ligation, amplification and library clean-up steps. There may also be size selection steps to enrich long fragments and remove shorter ones that may affect sequencing accuracy.

Tip 4: Multiplexing for simultaneous sequencing of multiple samples

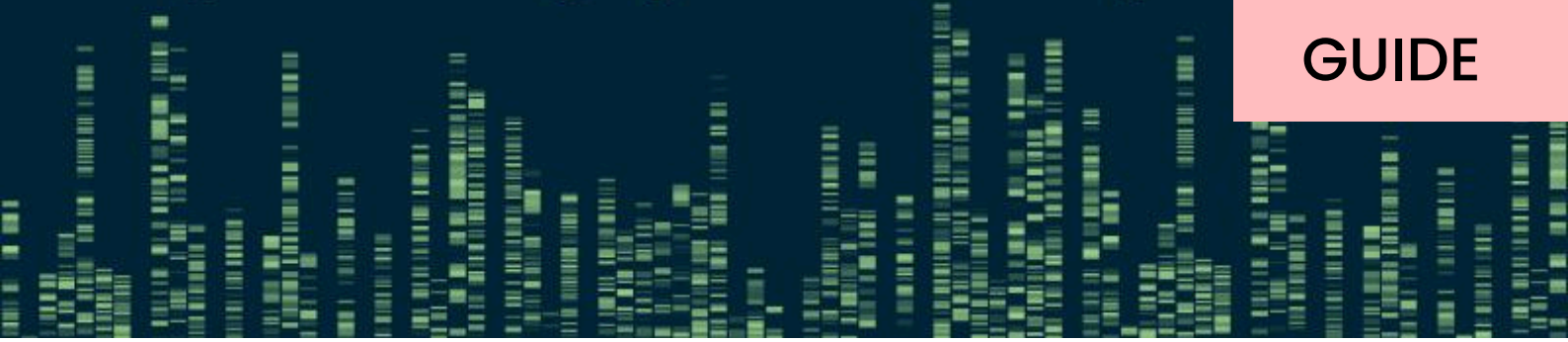
Some library preparation kits come with a selection of unique barcodes, which allow multiple samples to be sequenced simultaneously, increasing throughput and cost effectiveness. The barcode ensures that you can identify and separate individual samples during downstream data analysis. If you do use multiplexing in your sequencing, use sample plate maps and multichannel pipettes to ensure samples are allocated the correct barcodes and throughput is increased. Also ensure samples are pooled to provide equal representation of each one, whilst remembering to consider desired data output and genome size if these are different between samples.

Tip 5: Library quality control

Before loading prepared libraries on the sequencing platform do a final library quality control check. You do not want to waste an expensive sequencing flowcell on a poor, failed library! Assess library quality using gel electrophoresis and fluorometric quantification, or similar methods, to confirm that library size distribution, quality and quantity are roughly correct. Monitor the performance of the sequencer using control samples or PhiX spike-ins to ensure optimal sequencing conditions.

Tip 6: Determine desired sequencing depth

Determine the appropriate sequencing depth for your sample. Consider factors such as genome size, complexity and the level of sensitivity required for variant detection, and consult published literature to help you determine an appropriate sequencing depth. Once you have reached the desired sequencing depth for your library, any data past this point might not be useful. Instead, you may be able to wash the current library off and then reuse the flowcell



for a different sample, to further increase cost effectiveness.

Tip 7: Bioinformatic analysis pipelines
Develop and adopt an appropriate, automated bioinformatics pipeline for data analysis. This could include tools to perform read quality assessment and filtering, assembly and polishing, assembly quality checks, variant calling and other downstream analyses to extract meaningful insights from the raw data. Remember to stay up to date with the latest versions of bioinformatics tools and methodologies relevant to your sequencing platform and analysis.

Tip 8: Data storage and backup
DNA sequencing generates vast amounts of data. Ensure you have adequate storage capacity and backup systems to handle and protect the generated data. Implement a reliable data management strategy to organize and store the sequencing data efficiently, facilitating data access and analysis.

Tip 9: Stay up to date
DNA sequencing and subsequent analysis are continuously evolving. Try to keep up to date with advancements in sequencing technologies, protocols and data analysis tools. Regularly review scientific literature, attend conferences and workshops, and engage with the scientific community to stay informed about the latest techniques and best practices.

Tip 10: Automate where possible
Many sequencing technologies have the capability to simultaneously sequence hundreds of samples at once. Now, the rate-limiting steps are generally in the preparation of these samples for sequencing and the subsequent data analysis. There are automated systems for DNA extraction, quality control and library preparation that can speed up the wet lab side of this. There are also ways to automate data backup and analysis via the command line or open, web-based platforms like Galaxy. You may even be able to outsource the entire process to external DNA sequencing companies who will take your extracted DNA (and money) and in return will give you processed sequencing data. This may be beneficial for some projects; however, you may generally have less control and will still have to consider the points covered above.

P.S. Final tip: never load a library at the end of the working day… Something always goes wrong, and it is always better to have friendly support around when this happens.

References

1. Fleischmann RD, Adams MD, White O *et al.* Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. *Science* 269(5223) 496–512 (1995).
2. Adams MD, Celniker SE, Holt RA *et al.*The genome sequence of *Drosophila melanogaster. Science* 287(5461), 2185–2195 (2000).
3. UK Health Security Agency. A revolution in genomic sequencing. [Date accessed: 11th July 2023].
4. Hayden EC. Genome sequencing: the third generation. *Nature* doi: 10.1038/news.2009.86 (2009) (Epub ahead of print).
5. Waters EV, Tucker LA, Ahmed JK, Wain J and Langridge GC. Impact of *Salmonella* genome rearrangement on gene expression. *Evol. Lett.* 6(6) 426–437 (2022).

About the Author

Emma Waters (right) is a Postdoctoral Research Scientist at the Quadram Institute (Norwich, UK). Emma originally trained as a Chemist at the University of East Anglia (Norwich, UK), before receiving her PhD in Biochemistry where she investigated how to attach tiny chemical solar panels to bacteria so they could harness solar energy to produce useful chemicals.

Since then, Emma changed her research field to medical microbiology where she uses different sequencing techniques to investigate bacterial niche adaptation. Currently, she uses long-read and RNA sequencing to identify and study the effects of large structural rearrangements in the genomic DNA of bacteria. This type of mutation shuffles genetic instructions to survive new stresses and environments. In the case of pathogenic bacteria, this can aid adaption to asymptomatic lifestyles, silent transmission and detection aviation.

Emma Waters is funded by Biotechnology and Biological Sciences Research Council (BBSRC) Institute Strategic Programme Microbes and Food Safety and has no identified conflicts of interest with the contents of this article.

BioTechniques

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/ibtn20

Long-Read Sequencing for the Metagenomic Analysis of Microbiomes

Tristan Free

To cite this article: Tristan Free (2023) Long-Read Sequencing for the Metagenomic Analysis of Microbiomes, BioTechniques, 74:4, 153-155, DOI: 10.2144/btn-2023-0028

To link to this article: https://doi.org/10.2144/btn-2023-0028

BioTechniques
The International Journal of Life Science Methods

LONG-READ SEQUENCING FOR THE METAGENOMIC ANALYSIS OF MICROBIOMES

One technology, long-read sequencing, and one research field, microbiome studies, have risen to prominence over the last decade. But how can one be used in the other? What changes are being wrought? And what limitations remain?

The rise of long-read sequencing over the last decade has been well documented, with the initial concerns surrounding accuracy being quickly addressed by the latest generation of technologies from Oxford Nanopore (Oxford, UK) and PacBio (CA, USA), which both produce sequences with a read accuracy greater than 99% (Figure 1) [1,2]. These developments have culminated in these techniques being named Nature's Method of the Year 2022 and have led to their spread into ever more complex applications and challenging research, perhaps best exemplified by their burgeoning use in the metagenomic analysis of microbiomes.

FALLING SHORT IN METAGENOMIC SEQUENCING

Metagenomic analyses aim to sequence and assemble the total genomes from all organisms present in a sample; an approach that proves nearly essential to the increasingly popular study of host and environmental microbiomes [3]. Traditionally, metagenomic sequencing has relied on short-read sequencing to provide the sequences with which to construct Metagenome Assembled Genomes (MAGs). However, these techniques have several limitations that present challenges that are difficult to overcome while still using short-read sequencing.

For instance, short-read sequencing often relies on PCR amplification of DNA in the sample, which often struggles to amplify GC-rich regions due to the increased stability of these DNA sequences, introducing GC bias into the data created. Furthermore, the length of short reads – typically between 75–400 bp – inhibits the sequencing of long repeat sequences and structural variants, limiting the ability to create MAGs with full sequence coverage [4].

Identifying specific strains of a bacterial species is another challenge that arises when you need to conduct a high-resolution analysis of your sample. Talking to Titus Brown, a metagenomic sequence analysis expert leading the Data Intensive Biology lab at the University of California, Davis (CA, USA), it becomes clear why this is so important in the study of microbiomes.

"Bacteria have dramatically more variation in their genomes than eukaryotes, which typically have small variations such as small nucleotide polymorphisms. Bacteria, meanwhile, will have the same core genome that composes roughly 60% of the genome but the remaining 40% will be a composite of different genes transferred from, and required to survive in, their specific environment."

It is highly challenging to obtain the linkage information required to confidently assemble a strain-specific genome using short-read sequencing. While attempts have been made to resolve this issue, it remains a serious roadblock to the use of the technique for this application [5]. Instead, short-read assembled MAGs often represent a composite genome of the strains of a species present in a sample. This loses a lot of the contextual information within the strain-specific genome that is needed to interpret the functions and interactions occurring in a microbiome.

ENTER LONG-READ SEQUENCING

While short reads present a serious challenge to assembling genomes from the same strain or cell, long reads are large enough to provide the linkage information to identify which specific microbial strain each read is from, enabling the assembly of strain-specific MAGs. Talking to *BioTechniques* for a recent Talking Techniques podcast episode, Jeremy Wilkinson of PacBio stated that, *"…one HiFi read has an average of eight intact genes per read, which allows for greater functional and taxonomic profiling, with 90% of the reads being annotatable or classifiable."* Therefore, if you are designing a study that intends to obtain specific information about the different strains of a species present in the genome, long-read sequencing techniques provide an excellent tool to deliver this [6].

For example, in a recent study led by the USDA Dairy Forage Research Center (WI, USA), Bickhart *et al.* used PacBio HiFi reads to generate 428 MAGs with over 90% completeness from sheep feces [6]. Developing their own software, MAGPhase, the team was able to identify 220 lineage-resolved MAGs, differentiating

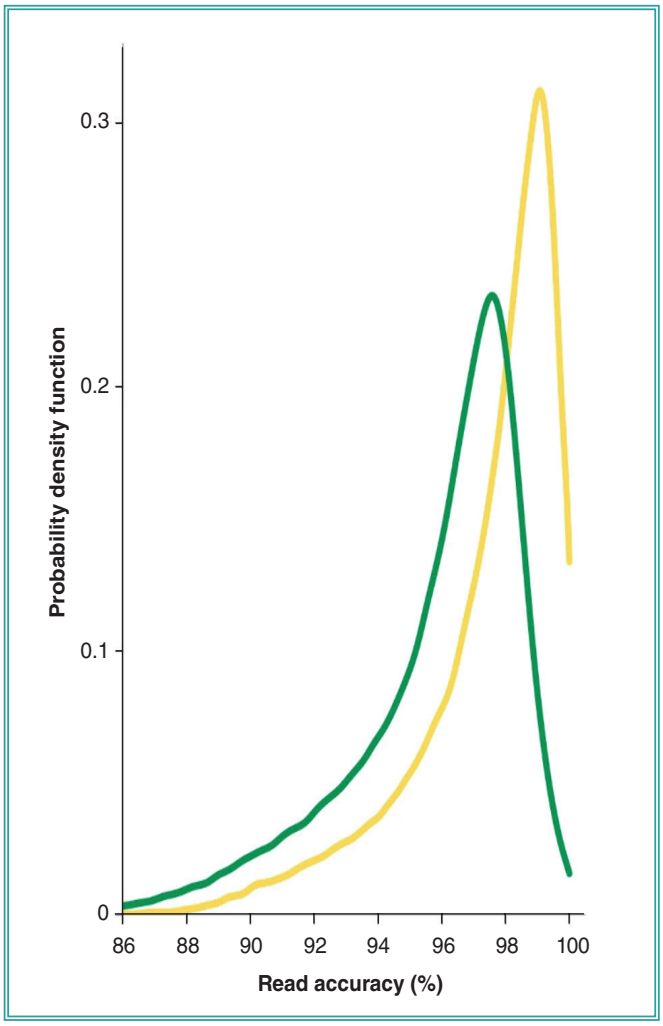


Figure 1. Observed raw read accuracies measured through read-mapping. Nanopore R.4.1 in green, Nanopore R10.4 in yellow. Figure adapted with permission from [1] under a Creative Commons Attribution 4.0 International License: http://creativecommons.org/licenses/by/4.0/

between related strains of the same species found within the fecal matter. With this additional contextual information, the team was able to identify 424 novel host-strain associations within the sheep fecal microbiome.

Similar metagenomic studies of canine fecal microbiomes have yielded interesting results, highlighting canid-specific bacteria and resolving key aspects of the mobilome, the mobile components of the genome that make up the composite part of bacterial genomes highlighted previously by Brown [8].

What's more, due to the exceptional size that long reads are now reaching, structural variants and long repeat sequences can now be captured within a single read, while the lack of reliance on PCR means that GC-rich sequences present a much less daunting challenge

REMAINING CHALLENGES

While long-read sequencing has resolved many issues in metagenomics, two closely linked challenges remain.

Firstly, As exemplified by the papers noted above, the majority of success stories of long-read sequencing for metagenomics have come from studies of host-associated microbiomes as opposed to environmental microbiomes [9]. This is because long-read sequencing requires higher quality and larger volumes of DNA compared to short-read sequencing. This challenge follows clear logic: to produce long reads, you need long, intact strands of DNA to sequence. In a host-associated microbiome, there is less diversity, meaning that each species' genome is present in a comparatively higher concentration than environmental samples and what's more, the DNA in these samples is also often in a much better condition.

This is compounded by the second challenge, the output of long-read technologies, which lags behind short-read sequencing, a technique that has had decades to optimize and increase the amount of sequence data that can be obtained from each sample. As a result, generating enough sequences to produce MAGs for each of the numerous species present in a complex environmental sample represents a current limitation of these techniques.

The ideal solution to this problem would be to establish a technique or methodology that combines long and short read sequencing techniques in an attempt to reach a perfect union between short read's ability to provide wide coverage of many, low abundance species, and long read's ability to resolve genomes of those species down to individual strains.

In the first steps towards this solution, studies utilizing long-read-assembled and hybrid-assembled MAGs to fill in the gaps left behind by short read-based metagenomic studies have borne fruit. For instance, a recent analysis of 109 gut microbiomes across three ethnicities in Singapore successfully compiled 1708 hybrid-assembled MAGs that had been missed by short-read-only analyses. This facilitated the discovery of 70 novel gut microbes and over 3400 strains, providing far greater insight into the gut microbiome of Southeast Asian populations [10].

As for the quantity and quality of DNA? Quantity poses a major challenge that is difficult to address. In an environmental sample, increasing the sample size to raise the volume of DNA that can be yielded from sample preparation only increases the diversity, stretching the output of these sequencing technologies. Ultimately, to resolve this issue the DNA input requirements for these technologies need to be reduced; a solution that is simple in principle but not in practice. Currently, Oxford Nanopore's technology has lower input requirements than PacBio's, however, this is balanced with slightly less accurate read calling.

Quality, on the other hand, has more readily available solutions. Sample preservation solutions exist that can stabilize the microbes in a sample and protect the nucleic acids present, such as the OMNIgene gut collection kit, and even immobilize pathogens in the sample to make it safe, like Zymo Research's DNA/RNA Shield medium.

IS IT ALL ABOUT THE ANALYSIS?

For sample input requirements and the sequencing output, time is the only solution; however, one that Brown is confident about. *"I'm not worried about the sequencing processes and the extraction methods. There is plenty of motivation out there to improve these and they will with time."* What concerns him is what follows after collection: analysis.

A growing trend in multiple fields of the life sciences is the acquisition of larger and more complex datasets, the assumption being that with more information, deeper and more meaningful insights will follow. Yet, in recent years, a counterargument has arisen: if the purpose of developments such as single-cell sequencing was to move away from bulk readings and averaged results, then why in other instances do we often see the generation of larger, deeper datasets as a tool to yield more perceptive insights? Whilst speaking at the 2022 iteration of the annual Society for Neuroscience conference (12–16 November, San Diego, CA, USA), Tim Harris (Howard Hughes Medical Institute, VA, USA) – the developer of Neuropixels, which enable the collection of signals from hundreds of individual neurons in different brain regions – posed the question of his own invention: *"Have Neuropixels led to studies that bury findings in vast swathes of data or have they enabled researchers to collect enough information to discover the emphatic truth?"* [11]

The key to this problem is designing data analysis techniques that can provide better insights than simply averaging datasets and grouping findings. This is what mostly concerns Brown. *"I think as a field we are mostly concerned with generating data, assuming that we will be able to do something useful with it at a later date. However, as it stands, I think there is a lot of room to improve data analysis."*

Fortunately, Brown is not alone in this concern. Significant efforts are underway to improve microbiome analysis, and there are many software platforms, such as Smash community, with which to conduct metagenomic data analysis in distinct contexts [12].

While data analysis does prove perhaps the biggest challenge facing the field, the current techniques available and ongoing research efforts are continuing to produce vitally important results that would not have been possible with previous technologies.

A fine example of this work includes the recent development of target-enriched long-read sequencing (TELSeq) by a research collaboration between the University of Florida (FL, USA) and the University of Minnesota (MN, USA). This technique was used to investigate antimicrobial resistance genes in three different metagenomes: human fecal microbiome transplant material, bovine fecal material from animals that had not received antibiotics and soil from unused prairie surrounded by farmland. This study revealed numerous colocalizations between mobile genetic elements and antibiotic-resistance genes in the soil and human fecal metagenomes. By demonstrating its ability to deliver results in both host-associated and environmental microbiomes, this study provides a significant indication of the vast strides being made in long-read sequencing and the value that long-read sequencing techniques can add to metagenomic analyses of the microbiome [13].

Written by Tristan Free

REFERENCES

1. Sereika M, Kirkegaard R, Karst S *et al.* Oxford Nanopore R10.4 long-read sequencing enables the generation of near-finished bacterial genomes from pure cultures and metagenomes without short-read or reference polishing. *Nat. Methods* 9(7), 823–826, (2023).
2. Hon T, Mars K, Young G *et al.* Highly accurate long-read HiFi sequencing data for five complex genomes. *Sci. Data* 7(1), 399 (2020).
3. National Human Genome Research Institute. Metagenomics. www.genome.gov/genetics-glossary/Metagenomics
4. Adewale B. Will long-read sequencing technologies replace short-read sequencing technologies in the next 10 years? *Afr. J. Lab. Med.* 9(1), 1340 (2020).
5. Beggel B, Neumann-Fraune M, Kaiser R, Verheyen J, Lengauer T. Inferring short-range linkage information from sequencing chromatograms. *PLOS ONE* 8(12), e81687 (2013).
6. Reiter T and Brown T. MAGs achieve lineage resolution. *Nat. Microbiol.* 7(2), 193–194 (2022).
7. Bickhart D, Kolmogorov M, Tseng K *et al.* Generating lineage-resolved, complete metagenome-assembled genomes from complex microbial communities. *Nat. Biotechnol.* 40(5), 711–719 (2022).
8. Cuscó A, Perez D, Viñes J, Fàbregas N, Francino O. Long-read metagenomics retrieves complete single-contig bacterial genomes from canine feces. *BMC Genom.* 22(1), 330 (2021).
9. Marx V. Method of the year: long-read sequencing. *Nat. Methods* 20(1), 6–11 (2023).
10. Counot JS, Chia M, Bertrand D et al. Genome-centric analysis of short and long read metagenomes reveals uncharacterized microbiome diversity in southeast Asians. *Nat. Commun.* 13(1), 6044 (2022).
11. *BioTechniques*. Talking Techniques. Neuropixels: big data heaven or burying the lead in averages? www.biotechniques.com/podcasts/talking-techniques-neuropixels-big-data-heaven-or-burying-the-lead-in-averages/
12. Navgire G, Goel N, Sawhney G et al. Analysis and interpretation of metagenomics data: an approach. *Biol. Proced. Online* 24(1), 18 (2022).
13. Slizovskiy I, Oliva M, Settle J et al. Target-enriched long-read sequencing (TELSeq) contextualizes antimicrobial resistance genes in metagenomes. *Microbiome* 10(1), 185 2022.

Epigenomics

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/iepi20

New insights into the dynamics of m6A epitranscriptome: hybrid-seq identifies novel mRNAs of the m6A writers METTL3/14

Konstantina Athanasopoulou, Panagiotis G. Adamopoulos & Andreas Scorilas

To cite this article: Konstantina Athanasopoulou, Panagiotis G. Adamopoulos & Andreas Scorilas (2024) New insights into the dynamics of m6A epitranscriptome: hybrid-seq identifies novel mRNAs of the m6A writers METTL3/14, Epigenomics, 16:17, 1159-1174, DOI: 10.1080/17501911.2024.2390818

To link to this article: https://doi.org/10.1080/17501911.2024.2390818

EPIGENOMICS
2024, VOL. 16, NO. 17, 1159–1174
https://doi.org/10.1080/17501911.2024.2390818

Taylor & Francis
Taylor & Francis Group

RESEARCH ARTICLE

New insights into the dynamics of m6A epitranscriptome: hybrid-seq identifies novel mRNAs of the m6A writers METTL3/14

Konstantina Athanasopoulou[a], Panagiotis G. Adamopoulos[a] and Andreas Scorilas[*,a]

[a]Department of Biochemistry & Molecular Biology, National & Kapodistrian University of Athens, Athens, 15701, Greece

ABSTRACT

Background: N6-methyladenosine (m6A), a prevalent mRNA modification, is dynamically regulated by methyltransferases, including METTL3 and METTL14.

Materials & methods: In the current study, we employed a custom hybrid-seq method to identify novel *METTL3/14* transcripts, explore their protein-coding capacities and predict the putative role of the METTL isoforms.

Results: Demultiplexing of the hybrid-seq barcoded datasets unraveled the expression patterns of the newly identified mRNAs in major malignancies as well as in non-malignant cells, providing a deeper understanding of the methylation pathways. Open reading frame query revealed novel METTL3/14 isoforms, broadening our perspective for the structural diversity within METTL family.

Conclusion: Our findings offer significant insights into the intricate transcriptional landscape of *METTL3/14*, shedding light on the regulatory mechanisms underlying methylation in mRNAs.

ARTICLE HISTORY
Received 26 March 2024
Accepted 5 August 2024

KEYWORDS
alternative splicing; methyltransferases; METTL14; METTL3; N6-Methyladenosine (m6A); nanopore sequencing; RNA modifications

1. Introduction

While epigenetic changes, such as DNA methylation and histone modification, are widely explored in molecular biology research, the study of mRNA modifications is an emerging field that encompasses various RNA methylations, such as the dominant N6-methyladenosine (m6A) and additional alterations including N1-methyladenosine (m1A), 5-methylcytosine (m5C) and 7-methylguanosine (m7G) [1]. Among these, m6A stands out as the most abundant epitranscriptomic mark, being deposited onto mRNAs by different methyltransferase proteins that act either independently or form multiprotein complexes [2]. Methyltransferase-like 3 (METTL3) introduces m6A individually or by forming stable complexes with Methyltransferase-like 14 (METTL14) and additional ligands, including the Wilms tumor 1-associated adaptor protein (WTAP), the Zinc-finger CCCH-type-containing 13 (ZC3H13), the Vir-like m6A methyltransferase-associated (VIRMA) and the RNA-binding motif protein 15/15B (RBM15/15B) [3–8]. Further studies suggest that Methyltransferase-like 16 (METTL16) also catalyzes the introduction of m6A at specific mRNA sites [9,10].

Multiple studies have extensively investigated the crystal structure and the function of the METTL3/METTL14 heterodimer, supporting the notion that this complex acts co-transcriptionally at the m6A consensus motif, DRACH (D = A, G, or U; R = G or A; and H = A, C or U) [11,12]. Regarding METTL3, the full-length protein comprises 580 amino acids, has a molecular weight of 70 kDa, and conserves distinct domains crucial for its methyltransferase activity. In brief, the N-terminus harbors the WTAP-binding domain (residues 1–36), which interacts with the splicing regulator WTAP to facilitate proper localization. METTL3 engages with mRNAs containing the DRACH motif through two CCCH-type zinc finger domains (ZFD), namely ZFD1 (residues 259–298) and ZFD2 (residues 299–336), forming an anti-parallel β-sheet [13,14]. At the C-terminus, the METTL3 methyltransferase domain (residues 357–580) is responsible for its enzymatic activity and features two additional motifs: the DPPW (residues 395–398) and the S-adenosylmethionine (SAM) binding site (residues 512–535). Both motifs are essential for transferring the methyl group to the adenine base. Specifically, the Asp of the DPPW domain is considered the catalytic residue for activating the N6 at the modified base, while the SAM binding site serves as the methyl donor [15].

On the other hand, the 53.3 kDa METTL14 includes 456 amino acids and harbors several catalytical domains [16]. The N-terminal of METTL14 conserves the coiled coil domain (residues 63–93) that is responsible for protein's interactions, while the methyltransferase domain (residues 165–378) is the predominant part of the protein. Moreover, the N-terminal α-helical motif that includes

CONTACT Andreas Scorilas Tel.: +30 2107274306; ascorilas@biol.uoa.gr

Supplemental data for this article can be accessed at https://doi.org/10.1080/17501911.2024.2390818

residues 116–163 and the C-terminal motif formed by residues 380–402, constitute the regions that enclose the methyltransferase domain. The methyltransferase domain (MTD) of METTL14 is similar to METLL3 and harbors the catalytic EPPL motif (residues 192–195) as well as the SAM binding site (residues 344–347) to transfer methyl groups to RNA. RGG/RG motif (residues 409–452) is an arginine-glycine rich domain at the C-terminal of METTL14 that facilitates the RNA binding process [14].

Undoubtedly, the protein structures and functionalities of both METTL3 and METTL14 have been extensively characterized [17]. However, up to date the transcriptional profile of both *METTL3* and *METTL14* has not undergone thorough investigation. According to GenBank's GRCh38 Annotation Release RS_2023_03, a total of four *METTL3* mRNAs have been characterized. These include the primary *METTL3* mRNA (*METTL3* v.1, accession number: NM_019852.5) and three predicted mRNA models (*METTL3* X1, X2 and X3 accession numbers: XM_011536968.3, XM_047431594.1 and XM_006720206.5, respectively). All the alternative transcripts are suggested to encode different protein isoforms, since their nucleotide sequence differs from *METTL3* v.1 transcript. On the contrary, *METTL14* is characterized by two mRNAs, the annotated *METTL14* v.1 (*METTL14* v.1, accession number: NM_020961.4) and the predicted X1 transcript (*METTL14* X1, accession number: XM_047416029.1). *METTL14* mRNAs consist of different transcriptomic regions thus encode different protein isoforms.

The primary objective of the current study is to delve into the transcriptional profile of *METTL3* and *METTL14* genes, aiming to identify novel mRNAs with potential protein-coding capacity. Additionally, our study seeks to discern the expression levels of the novel *METTL3* and *METTL14* splice variants in multiple human cancers and non-cancerous cell lines. To achieve this, we designed a hybrid-seq approach involving multiplex nanopore sequencing coupled with short-read NGS sequencing, to generate 'polished' full-length nanopore reads and identify novel mRNA transcripts. In this study, we emphasize the presence of alternative *METTL3* and *METTL14* mRNA transcripts and highlight their expression levels across various human malignancies. Furthermore, we conducted bioinformatics analysis to predict the protein-coding capacities of the newly described mRNAs and compared our findings with results from Translating Ribosome Affinity Purification sequencing (TRAP-Seq) and Poly-Ribo-Seq applications. Finally, our study discusses the putative biological functions of the predicted METTL3 and METTL14 isoforms, providing valuable insights into the intricate transcriptional landscape of these major methyltransferase genes and shedding light on m6A methylation in human mRNAs.

2. Materials & methods

2.1. Human cell line culture & total RNA isolation

The present work was carried out in a wide panel of human cancer cell lines originating from breast cancer (MDA-MB-468, MDA-MB-281, MCF-7, BT-474, SKBR-3 and BT-20), liver cancer (HepG2 and Huh-7), ovarian cancer (OVCAR-3 and ES-2), prostate cancer (PC-3 and LNCaP), gastric cancer (AGS), colorectal cancer (Caco-2, DLD-1, HT-29, HCT 116, SW 620) and hematological malignancies (HL-60, THL-1, U937, Raji and SUDHL1). Additionally, we utilized the non-cancerous human embryonic kidney (HEK-293) and human keratinocytes (HaCaT) cells. All cell lines were cultured based on the American Type Culture Collection guidelines.

Total RNA was isolated from each human cell line using the TRIzol Reagent (Ambion™, Thermo Fisher Scientific Inc., Waltham, MA, USA) and all the RNA samples were appropriately diluted in THE RNA Storage Solution (Ambion™). The assessment of their purity and concentration was tested spectrophotometrically at 260 and 280 nm, using BioSpec-nano Micro-volume UV–Vis Spectrophotometer (Shimadju, Kyoto, Japan).

2.2. mRNA enrichment & first-strand cDNA synthesis

The positive selection of poly(A) RNAs was accomplished using the NEBNext® Poly(A) mRNA Magnetic Isolation Module (New England Biolabs, Inc., Ipswich, MA, USA) and an initial RNA amount of 5 μg. The poly(A)+ RNA samples were quantified with the Qubit® RNA HS Assay Kit (Invitrogen™, Thermo Fisher Scientific Inc.). Accordingly, reverse transcription assays were performed to generate first strand cDNAs. Briefly, the initial reaction mixtures included ~80–100 ng of poly(A) RNA from each human cell line and 1 μl oligo-dT_{20} (10 μM), were incubated at 65°C for 5 min in a Veriti 96-Well Fast Thermal Cycler (Applied Biosystems™) and afterward were placed on ice for 2 min. Then each reaction was completed by adding 1 μl dNTP mix (10 mM each), 0.5 μl (20 U) RNaseOUT inhibitor and 1 μl (200 U) of the reverse transcriptase Maxima H Minus (Invitrogen™, Thermo Fisher Scientific Inc.). The cDNA synthesis was carried out by incubating the reaction mixtures at 50°C for 40 min and subsequently the enzyme was deactivated by heating the mixtures at 85°C for 5 min. The human housekeeping *GAPDH* (Glyceraldehyde 3-phosphate dehydrogenase) gene was amplified for the quality assessment of the cDNA. The cDNA samples were then mixed equimolarly to generate

one cDNA pool sample per type of malignancy (breast, liver, ovarian, hematological, prostate, gastric and colorectal). Finally, one cDNA pool sample was created from the non-cancerous cell lines HEK-293 and HaCaT.

2.3. Primer design & PCR amplification

For each *METTL* target gene, a PCR-based assay using two gene-specific primers (GSPs) was designed and employed for the specific amplification of their mRNAs. Specifically, for *METTL3* a forward GSP (F: 5′-TCAGGGCTGGGAGACTAGGA-3′) targeted the first annotated exon and was used along with a reverse GSP (R: 5′-GGATTCTTAGCTCTGTAAGGAAGTGC-3′) that was designed to anneal close to the annotated stop codon at the last exon of the *METTL3*. As for *METTL14*, the forward GSP anneals near the ATG site, the reverse targets the 11th annotated exon of the gene and their nucleotide sequences are 5′-TGGAACATGGATAGCCGCTTG-3′ and 5′-AGAAGGTTAGAGGAGGATGAATAGG-3′, respectively.

PCR assays were performed for the amplification of both *METTL3* and *METTL14* transcripts using the generated cDNA pools. Each PCR was carried out in reaction volumes of 25 μL with 0.2 mM dNTPs, 10X KAPA Taq Buffer A (Kapa Biosystems Inc.), which included $MgCl_2$ at a final concentration of 1.5 mM, 0.4 μM of each primer, and 1 U of KAPA Taq DNA Polymerase (Kapa Biosystems Inc.). The mix was amplified in a hot-lid Veriti 96-Well Fast Thermal Cycler under the following thermal conditions: 95°C for 3 min, and 35 cycles of 95°C for 30 s, 60°C for 30 s and 72°C for 3 min. A final extension step of 72°C for 5 min was also applied. Amplified products were electrophoresed on a 2% Agarose gel with EtBr stain at 100 V for the assessment of the PCR specificity. Finally, both METTL products from each cDNA pool were purified using the NucleoSpin® Gel and PCR Clean-up kit (Macherey-Nagel GmbH & Co. KG, Duren, Germany) and the amplicons were quantified with the Qubit® dsDNA BR Assay Kit (Invitrogen™, Thermo Fisher Scientific Inc.).

2.4. Targeted nanopore sequencing

An initial amount of 1 μg PCR product was used for the construction of barcoded DNA-seq libraries using the Ligation Sequencing Kit (SQK-LSK109, ONT) and the Native Barcoding Expansion Kit 1–12 (EXP-NBD104, ONT), according to the guidelines of the manufacturer. Briefly, end repair was accomplished with NEBNext® Ultra™ II End Repair/dA-Tailing Module (New England Biolabs, Inc), the adapter ligation step was completed utilizing the Quick T4 Ligase (New England Biolabs, Inc), while clean-up workflows were performed with the Agencourt AMPure XP beads (Beckman Coulter, Brea, CA, USA). The purified barcoded libraries were quantified with Qubit® dsDNA BR Assay Kit (Invitrogen™, Thermo Fisher Scientific Inc.) and the final library mix included all barcoded libraries, which were mixed equimolarly. Nanopore sequencing was conducted on a MinION Mk1C sequencer (Oxford Nanopore Technologies Ltd, ONT, Oxford, UK), using a FLO-MIN106D flow cell with R9.4.1 chemistry.

2.5. Next-generation sequencing

To perform polishing of the long sequencing reads, the same PCR product samples that were sequenced with nanopore sequencing were also used as templates for an NGS DNA-seq run using the semiconductor sequencing technology. The DNA-seq library preparation steps included enzymatic fragmentation, adapter ligation, nick-repair and purification of the ligated DNA according to the manufacturer's protocol and was completed with the Ion Xpress™ Plus Fragment Library Kit (Ion Torrent™, Thermo Fisher Scientific Inc.) To enrich the library for 300–400 bp fragments, bead-based size selection was carried out using the KAPA Pure Beads (Kapa Biosystems Inc.). Accordingly, the quantification of the NGS library was performed on a QuantStudio™ 5 Real-Time PCR System (Applied Biosystems™) with the Ion Library TaqMan™ Quantitation Kit (Ion Torrent™). The pre-sequencing template preparation and enrichment steps were performed using the Ion PGM™ Hi-Q™ View OT2 kit (Ion Torrent™), while the Ion PGM™ Hi-Q™ View Sequencing kit was used for the sequencing procedure on the Ion Personal Genome Machine™ (PGM™) platform.

Additionally, a typical RNA-seq experiment was carried out to compare the results not only with our designed nanopore sequencing approach, but also with findings from publicly available datasets from TRAP-Seq and Poly-Ribo-Seq applications. Our RNA-seq run was carried out using 10 ng poly(A)+ RNA sample from HEK-293 cells, while the library was prepared following the protocol of MGIEasy RNA Directional Library Prep Set kit (MGI Tech Co., Ltd). In the next step, the NGS run was carried out at a DNBSEQ-G50 (MGI Tech Co., Ltd) platform, using a FCL flow cell and PE100 read length. Finally, for comparison purposes, we utilized two publicly available sequencing datasets from Sequence Read Archive database of NCBI, which included a TRAP-Seq approach on human pluripotent stem cells (Study ID: PRJNA687244) and a Poly-Ribo-Seq approach on HEK-293T cells (Study ID: PRJNA853906).

2.6. Computational analysis of the sequencing data

The raw nanopore sequencing data was basecalled and demultiplexed with Guppy, which categorized sequencing reads to pass and fail in terms of quality. Only the sequencing reads that met the default quality cut-off

were taken into consideration for further analysis, while failed reads were discarded. Nanopore sequencing reads were then "polished" using specialized hybrid error correction algorithms under the recommended parameters. Polished long-read sequencing reads were then aligned to the human reference genome (GRCh38) by minimap2 aligner [18]. The derived BAM files that were generated were visualized with Integrative Genomics Viewer (IGV) for an initial assessment of the results. Additionally, the short sequencing reads produced by NGS were aligned to GRCh38 with hisat2 aligner under the default parameters.

2.7. Differential expression analysis of the identified METTL mRNAs

It should be noted that the multiplexing options of nanopore sequencing enabled the quantification of each alternative transcript in the investigated human cell lines. The *in-house* developed Alternative Splicing Detection Tool, namely "ASDT," was primarily used for the identification of alternative *METTL3* and *METTL14* transcripts [19]. "ASDT" is a specialized algorithm, written in PERL programming language, developed to identify and analyze alternative splicing events, including exon skipping events and intron retentions from massive parallel sequencing data. ASDT requires raw datasets (fastq files) as well as the reference annotation Genbank® file for any gene of interest to perform splicing analysis. The tool is designed to extract all exon coordinates, generate kmers for every possible alternative splicing event between all annotated exons of the gene, and then provide the abundance of each splicing event. Additionally, it offers the user the option to detect splicing sites allowing a number of mismatches, which is defined by the user. In the present study, only perfect matches were taken into consideration since the raw sequencing data were "polished" long-read sequencing reads. To estimate the expression levels for each *METTL3* and *METTL14* transcript, the reads per million (RPM) method was employed. Heatmaps were designed for visualizing the differential expression levels of the newly identified mRNAs.

2.8. Quantitative real-time PCR

The relative mRNA abundance of the described *METTL* mRNAs was investigated by RT-qPCR assays, using variant-specific primers, which were designed to target the novel splice junctions of *METTL3* and *METTL14* genes (Supplementary Tables S1 & S2). The qPCR-based assays were performed using cDNAs from HEK-293, MCF-7, OVCAR-3 and PC-3 cells as templates. All qPCR reactions were implemented in a QuantStudio™ 5 Real-Time PCR System (Applied Biosystems™). Amplification of each alternative mRNA was carried out in reaction volumes of 10 μL that included 5 μL of the 2X Kapa SYBR Fast qPCR Master Mix (KapaBiosystems, Inc.), 2 μM of each primer, and 1 μL of the cDNA template, while the thermal protocol included an initial denaturation step at 95°C for 3 min and 40 cycles of 95°C for 3 s and 60°C for 30 s. The human *GAPDH* housekeeping gene was also amplified and used as the endogenous reference control for normalization.

2.9. Statistical analysis

For qPCR data analysis, the statistical analysis software GraphPad Prism 9 (GraphPad Software, San Diego, CA, USA) was used. All statistical analysis results are presented with error bars representing mean ± standard deviation (SD). Detailed statistical tests performed for each *METTL* transcript comparing the results of the three biological replicates produced through qPCR.

2.10. Bioinformatics investigation of the coding capacities of the METTL3 & METTL14 mRNAs

To predict the coding capacity of novel mRNAs we utilized the ExPASy translate tool, which is designed to provide all putative initiation and termination codons of a given nucleotide sequence. To further investigate the putative role of the coding *METTL3* and *METTL14* transcripts, the predicted amino acid sequences were tested for the presence of the METTL conserved motifs. DeepGOWeb Protein Function prediction webserver provided the Gene Ontology (GO) enrichment analysis including confidence scores for the putative proteins [20]. Finally, AlphaFold3 was utilized to predict and generate 3D structures of the putative METTL isoforms, providing further insights into their structural characteristics and potential functional implications [21]. Additionally, functional insights based on 3D protein structures were elucidated using PyMOL software, which facilitated alignment and comparison of predicted isoforms with annotated METTL structures. In addition, RMSD (Root Mean Square Deviation) was calculated to assess structural similarities and differences. Moreover, the DeepFRI tool employed to assess high confident functional predictions of these proteins based on their 3D structures [22]. DeepFRI utilizes structural data from pdb files to annotate functional roles based on the GO terms for Cellular Component, Molecular Function and Biological Process.

3. Results

3.1. Nanopore sequencing unveils novel alternative METTL3 & METTL14 mRNAs

In the present study, we performed multiplexed amplicon sequencing using as templates cDNA pools originating from seven human cancers including breast (MDA-MB-468, MDA-MB-281, MCF-7, BT-474, SKBR-3 and BT-20), liver (HepG2 and Huh-7), ovarian (OVCAR-3 and ES-2), prostate (PC-3 and LNCaP), gastric (AGS), colorectal (Caco-2, DLD-1, HT-29, HCT 116 and SW 620) and hematological (HL-60, THL-1, U937, Raji, SUDHL1) malignancies as well as non-cancerous cells (HEK-293 and HaCaT). For each cDNA pool sample, two distinct amplicons were generated, corresponding to amplified mRNAs of METTL3 & 14. Bioinformatics analysis revealed a plethora of splicing events including exon skipping and intron retention, generating various alternative mRNAs, which have also been observed in the full length aligned reads visualized by IGV. Additionally, the coding capacities of the newly identified *METTL3* and *METTL14* transcripts were evaluated by the presence of both ORFs in their nucleotide sequence and conserved motifs in the generated amino acid sequence. On the contrary, RNAs harboring a premature termination codon (PTC) residing >50 nt upstream of the last exon junction were considered non-coding RNAs and candidates for nonsense-mediated mRNA decay (NMD) pathway [23].

Nanopore sequencing confirmed the existence of the annotated *METTL3* v.1 (NM_019852.5) and the previously predicted mRNA, *METTL3* X.1 (XM_011536968.3) but also a total of eight novel mRNAs was unraveled (*METTL3* v.2 - v.9). Briefly, seven mRNAs are generated by alternative splice junctions between the annotated exons (*METTL3* v.2 - v.8), whereas *METTL3* v.9 is characterized by the retention of two intronic regions between exons 8 and 9 and exons 9 and 10 (Figure 1). Our analysis confirms an alternative splice junction between exons 1 and 4, generating the alternative transcript variant *METTL3* v.2, which is characterized as non-coding RNA. Although *METTL3* v.3 is characterized by the entire absence of exons 6, 7 and 8, its mRNA sequence harbors the annotated start and stop codons and an ORF of 468 aa, hence is expected to be protein-coding (Figure 1). On the contrary, *METTL3* v.4 lacks exons 4, 5 and 6, contains a PTC and therefore is most likely a non-coding RNA. Exon skipping events occurring in *METTL3* v.5, generate an alternative splice junction between exon 3 and exon 10. *In silico* analysis revealed that although the entire lack of six annotated exons, *METTL3* v.5 is a coding mRNA since it harbors the annotated initial and termination codons (Figure 1). Exclusion of the annotated exon 6 generates an additional mRNA, *METTL3* v.6, which represents a non-coding RNA. Accordingly, *METTL3* v.7 displays a different splice junction of exon 5 with exon 10, while *METTL3* v.8 is formed by all the annotated exons except from exon 7. The applied ORF analysis showed that both *METTL3* v.7 and *METTL3* v.8 are most likely protein-coding since they possess ORFs from the annotated ATG the termination codon (Figure 1). Finally, *METTL3* v.9, which constitutes the only transcript harboring intronic sequences, is a non-coding RNA.

Similarly, our long-read sequencing approach unraveled the presence of eight alternative *METTL14* mRNAs (*METTL14* v.2–v.9). Bioinformatics analysis revealed that the lack of exon 2 in *METTL14* v.2, as well as the absence of exon 7 in *METTL14* v.3, generate non-coding RNAs candidates for NMD. *METTL14* v.4 lacks exon 8, but includes the annotated start and stop codons and possesses an ORF of 425 aa (Figure 2). Additionally, the alternative splice junction between exons 9 and 11 in *METTL14* v.5 introduces a novel stop codon located on the last exon, thus the mRNA is likely coding. Moreover, our approach failed to detect the predicted *METTL14* X1 transcript, which has an alternative exon 10 and lacks the entire exon 11, due to the specificity of the reverse primer which anneals at the last annotated exon. However, *METTL14* X1 harbors an additional exon between the annotated exons 8 and 9, namely C, which is also present in the newly described *METTL14* v.6 and v.7 mRNAs. Although *METTL14* v.7 includes the cryptic exon C, it lacks exon 10, whereas *METTL14* v.6 includes all the known exons and the C exon as well (Figure 2). Both transcripts lack ORFs and are highly expected to be non-coding. The last two mRNAs, harbor a novel cryptic exon between the exons 8 and 9, namely N. *METTL14* v.8 comprises of the 11 annotated exons and the N exon and represents a non-coding variant, whereas *METTL14* v.9, which includes the exon N but lacks the exon 10 possesses an ORF of 291 aa and shares the same annotated initiation and stop codons (Supplementary Figure S1).

3.2. Expression profiling of the METTL3 & METTL14 transcript variants

To conduct expression profiling of *METTL3* and *METTL14* mRNAs, long-read sequencing datasets were demultiplexed and RPM method was employed to assess the relative expression levels of each described *METTL* transcript among the investigated human cell lines. As expected, analysis showed that *METTL3* v.1 represents the most prevalent mRNA in all the barcoded libraries, whereas the rest mRNAs demonstrated diverse expression levels. Surprisingly, compared with all the alternative *METTL3* mRNAs, *METTL3* v.9 exhibits significantly higher expression levels across all the seven human malignancies

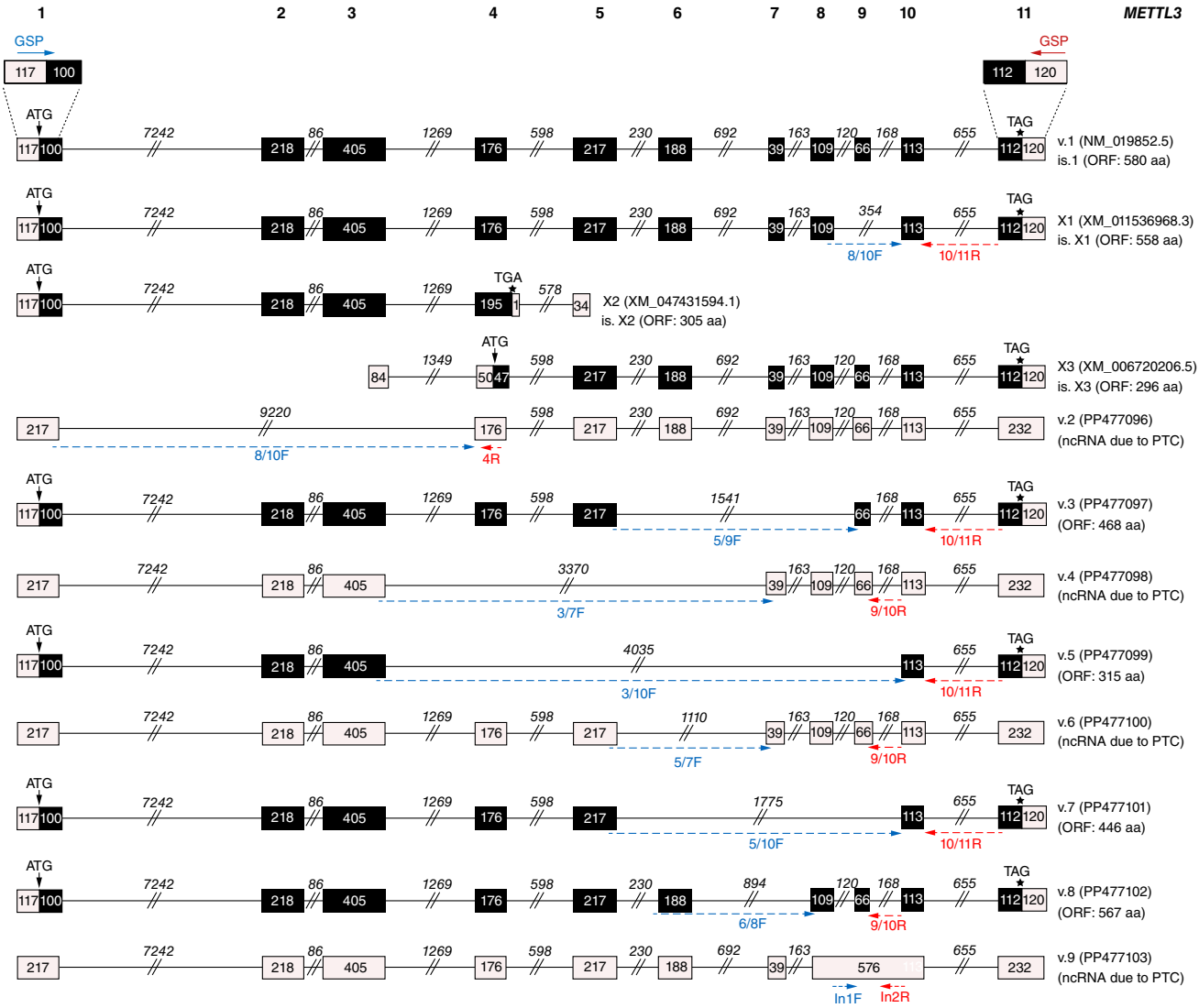


Figure 1. Structural illustration of the novel *METTL3* mRNAs (*METTL3* v.2–v.9). Exons are shown as boxes and introns as lines; black boxes demonstrate the coding regions of the mRNAs that harbor ORFs, while light red boxes represent non-coding regions. Non-coding RNAs that have a premature termination codon are also depicted in light red. Numbers inside boxes and above lines indicate the length of each exon and intron in nucleotides, respectively. Arrows define the position of the initiation codon (ATG), while asterisks (*) the position of the termination codon. The transcript number, the ORF length (only for protein-coding transcripts), and the GenBank® accession number are demonstrated next to each transcript. Finally, blue and red arrows represent relative position of the forward and reverse primers used for mRNA amplification.

and the non-cancerous cell lines, indicating its putative role into cells (Figure 3A). In addition, *METTL3* X1 is mainly expressed in breast and gastric cancer, and in hematological malignant cells, whereas *METTL3* v.4 is detected in higher levels in breast cancer and hematological malignant cells but also in non-cancerous cells. Notably, *METTL3* v.2 and v.6 are represented by higher RPM values in colorectal and gastric cancer, respectively. As for the remaining *METTL3* mRNAs, they were identified in significantly lower read counts as compared with v.1 and v.9 (Figure 3A).

Similarly, the main *METTL14* transcript is represented by higher RPM values in all the investigated human cell lines. Among the identified *METTL14* variants, the coding *METTL14* v.5 transcript emerged as the most abundant across all the barcoded libraries, hence it's highly promising to possess critical functional roles (Figure 3B). Expression analysis revealed that *METTL14* v.7 abundance displays a wide range of differentiation among the panel of human cell lines, being sufficiently expressed in ovarian, prostate and colorectal cancer but rarely represented in the remaining cell lines (Figure 3B). In addition, the

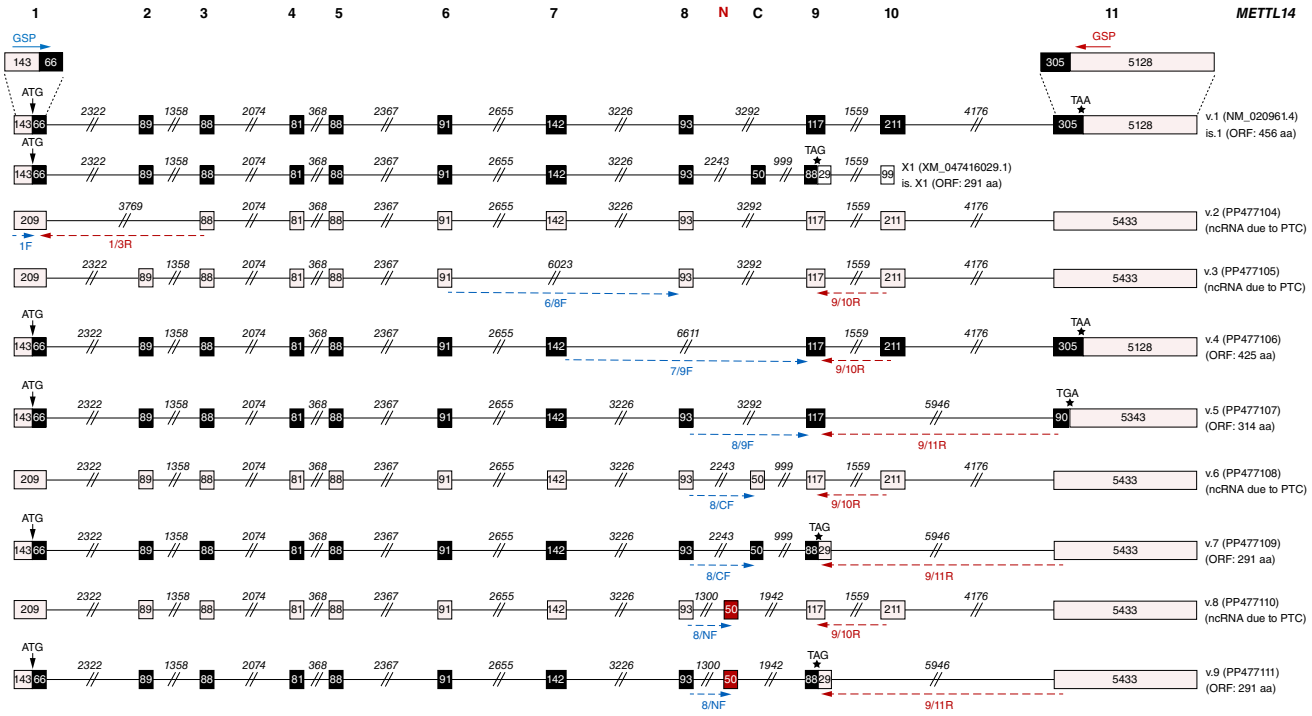


Figure 2. Schematic illustration of the novel *METTL14* splice variants (*METTL14* v.2–v.9). Exons are shown as boxes and introns as lines. The numbers inside the boxes and above the lines represent the length of each exon and intron, accordingly. Black boxes denote the coding region of the main *METTL14* v.1. Light red boxes are used to indicate the non-coding regions of the splice variants and the non-coding RNAs with PTCs. Finally, blue and red arrows represent relative position of the forward and reverse primers used for mRNA amplification.

remaining transcripts are significantly underrepresented as compared with *METTL14* v.1 and v.5, as they are hardly detectable among all the human malignancies (Figure 3B). Besides the comparative expression analysis of the described *METTL* transcripts among the investigated barcoded libraries based on their RPM values, complete clustering analysis (bidirectional cluster orientation) of transcript expression across the investigated human cells revealed distinct expression patterns for each alternative *METTL3* and *METTL14* variant (Supplementary Figure S2).

In addition to long-read sequencing data, we utilized RNA-seq to further enrich our analysis of alternative splicing events in *METTL3* and *METTL14* genes. Given the limitations of RNA-seq in capturing full-length mRNA sequences, our analysis focused on detecting splice junctions. From the RNA-seq data, we successfully identified all annotated splice junctions comprising the main *METTL3* transcript, as well as the two intronic sequences that comprise *METTL3* v.9 (Figure 3C). It's worth mentioning that the expression levels of intron retentions representing *METTL3* v.9 were notably elevated, a finding that is consistent with the obtained evidence from nanopore sequencing data. This concordance underscores the reliability of our observations across different sequencing platforms and reinforces the significance of *METTL3* v.9 in cellular processes (Figure 3C). RNA-seq also confirmed the existence of less abundant splice junctions, such as the junction between exons 3 and 7, the junctions of exon 5 with exon 7 and 10 as well as the splicing event between exons 6 and 8, which characterize the alternative *METTL3* v.4, v.6, v.7 and v.8 transcripts, accordingly (Figure 3C). In the case of *METTL14*, RNA-seq revealed the detection of the exon skipping event of exon 10 (Figure 3D). Despite the limited depth of coverage inherent to RNA-seq, the prevalence of the 9/11 junction that comprises *METTL14* v.5, supports the notion that v.5 is the most abundant previously uncharacterized *METTL14* variant. Our findings further underscore the importance of integrating multiple sequencing approaches to comprehensively elucidate the landscape of alternative splicing in mRNA methylation machinery.

3.3. mRNA expression analysis through quantitative PCR assays

Quantitative PCR-based assays revealed that the described *METTL3* and *METTL14* mRNAs can be successfully quantified in multiple human cell lines, being detected in all cell lines that were used (Figure 4).

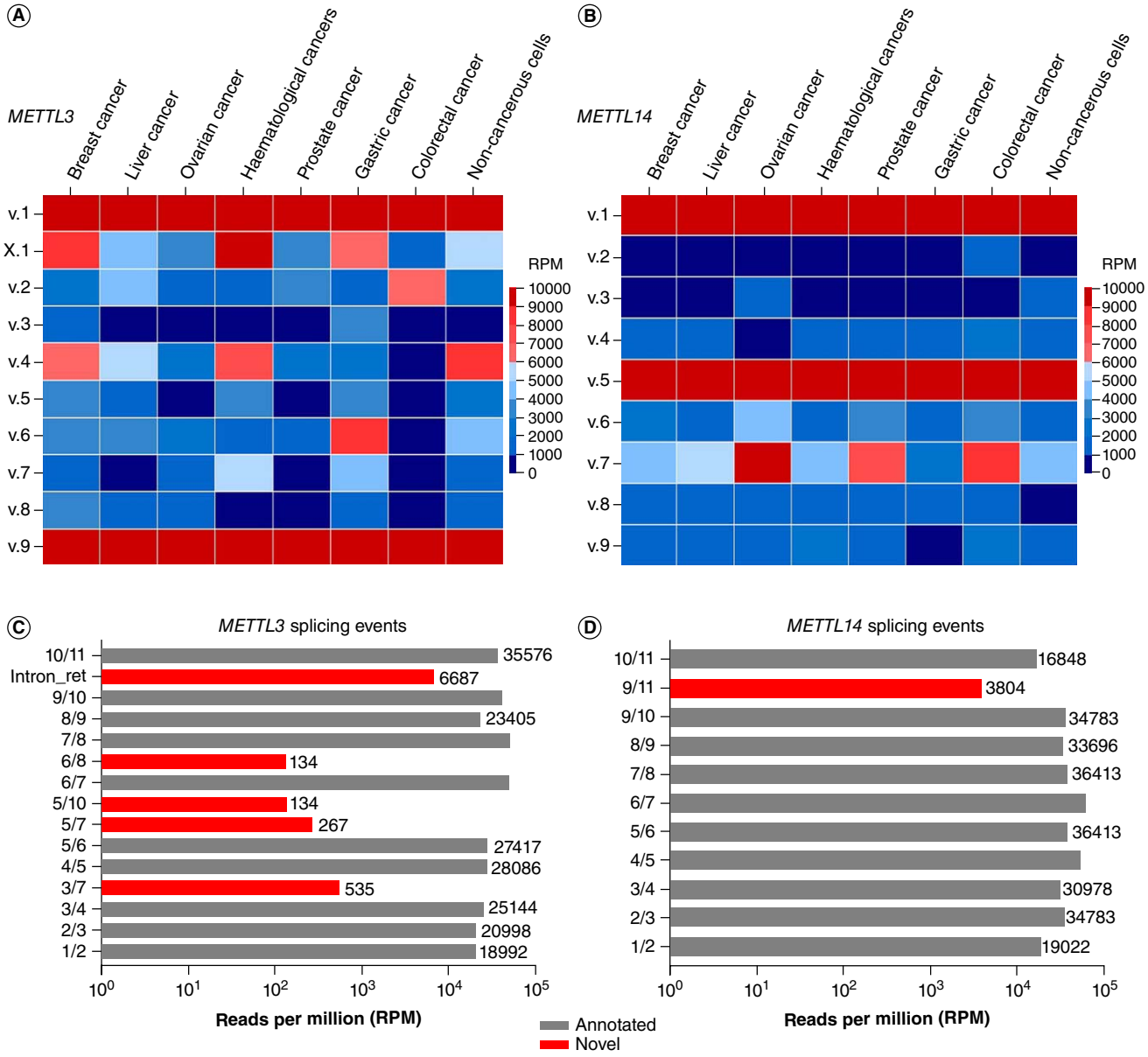


Figure 3. Relative expression of the novel *METTL* transcript variants and the detected splicing events of the present study. **(A)** Heatmap showing the expression levels of *METTL3* transcripts relative to the abundance of the annotated mRNA (*METTL3* v.1). **(B)** Heatmap illustrating the relative abundance of the alternative *METTL14* transcripts as compared with the annotated mRNA (*METTL14* v.1). **(C)** Read per million values of the annotated and novel splicing events of *METTL3* gene from NGS RNA-seq run in HEK-293 cells. **(D)** Read per million values of the annotated and novel splicing events of *METTL14* gene from NGS RNA-seq run in HEK-293 cells.

Each qPCR assay was conducted in triplicates to ensure robustness and reliability of the observed results.Of note, METTL3 v.9 demonstrated notable expression levels, being in accordance with our *in silico* analysis from the two distinct sequencing approaches utilized (Figure 4A). Moreover, *METTL3* X.1 is among the most abundant mRNA along with v.9 in cancer cells and in HEK-293. As for *METTL14*, most of the mRNAs are detected in the human cell lines tested; however their expression levels are differentiated (Figure 4B). It should be noted that *METTL14* v.5, which has been proposed as the most abundant alternative transcript according to nanopore sequencing reads, demonstrated noteworthy expression levels as compared with the rest identified mRNAs (Figure 4B). Additionally, *METTL14* v.7 is also detected in high levels. Although *METTL14* v.2 has been hardly detected through nanopore sequencing, qPCR assays revealed its high abundance in the human cell lines tested.

3.4. Translatome profiling supports the existence of novel protein-coding METTL3/14 mRNAs

To further decipher the protein-coding capacities of the identified *METTL* transcripts, two available translatome

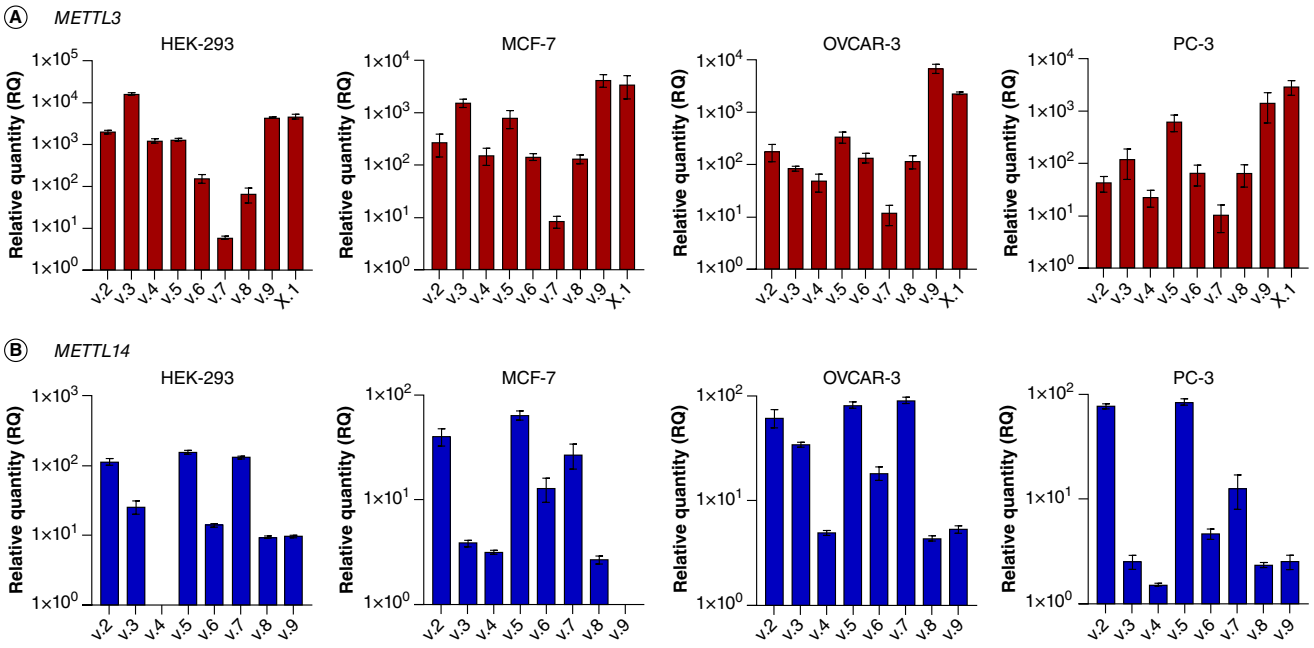


Figure 4. Barplots showing the relative abundance of each described transcript variant using qPCR-based assays and variant-specific primer pairs for **(A)** *METTL3* and **(B)** *METTL14*. The *GAPDH* housekeeping gene was utilized for normalization. The relative quantity of the described METTL3/14 mRNAs is exhibited as mRNA copies/10^6 GAPDH copies. All qPCR assays were performed using three biological repeats.

profiling studies (TRAP-Seq and Poly-Ribo-Seq) from Sequence Read Archive database were investigated for the potential detection of the described *METTL* transcripts. Analysis of Poly-Ribo-Seq dataset in HEK-293T cells confirmed the intron retention events that comprise *METTL3* v.9 (Supplementary Figure S3). Additionally, the exon skipping event between exons 5 and 7, which is included in v.6, as well as the splice junction between exon 3 and 10, which characterizes *METTL3* v.5 were also detected (Supplementary Figure S4A). Furthermore, TRAP-Seq data contained splicing events characterizing *METTL3* v.9 and v.6, but also the splice junction between exons 6 and 8 that characterizes *METTL3* v.8 (Supplementary Figure S4B). Additionally, analysis of the Ribo-seq peaks revealed which exons are translated into proteins. Hence, for the newly identified alternative transcripts, Ribo-seq peaks showed that the aforementioned alternatively spliced junctions are captured by ribosomes, providing strong evidence for the translation potential of these transcripts. *METTL3* v.9 exhibits strong Ribo-seq peaks at the intron retention region, suggesting that the transcript is actively translated into protein. Similarly, Ribo-seq signal intensities for exon skipping events characterizing *METTL3* v.5, v.6, and v.8 indicate their translation activities (Figure 5).

It should be noted that according to the obtained ORF query analysis, *METTL3* v.6 and v.9 have been classified as lncRNAs, however, sequencing reads from Poly-Ribo-Seq experiments suggest that these transcripts may constitute mRNAs with translation potential. This discrepancy can be explained since our analysis has been limited to detect coding mRNAs utilizing the annotated ATG and hence *METTL3* v.6 and v.9 might enclose alternative start codon(s). Additionally, both *METTL3* v.6 and v.9 harbor a PTC, which in some cases may be suppressed during translation. In the same manner, both TRAP-Seq and Poly-Ribo-Seq confirmed the presence of splicing event between exons 9 and 11, supporting the notion that *METTL14* v.5, v.7 and v.9 encode METTL14 isoforms (Supplementary Figure S4). TRAP-Seq has also revealed the coding potentials of *METTL14* v.3, although our *in silico* analysis suggests that v.3 represents a lncRNA. Moreover, Ribo-seq signal intensities for *METTL14*, highlight that the peaks for both splice junctions 9/11 and 6/8 correspond to ribosome-bound regions, indicating active translation sites (Figure 5).

3.5. *In silico analysis of the putative METTL3 & METTL14 protein isoforms*

Both METTL3 and METTL14 possess conserved domains, significant for their cellular functionality (Supplementary Figures S5 & S6). To investigate the structural integrity and potential functionality of these domains in the novel isoforms, *in silico* ORF query, AlphaFold3 structural predictions, PyMOL visualizations, and DeepFRI functional analyses were performed. Surprisingly, the highly

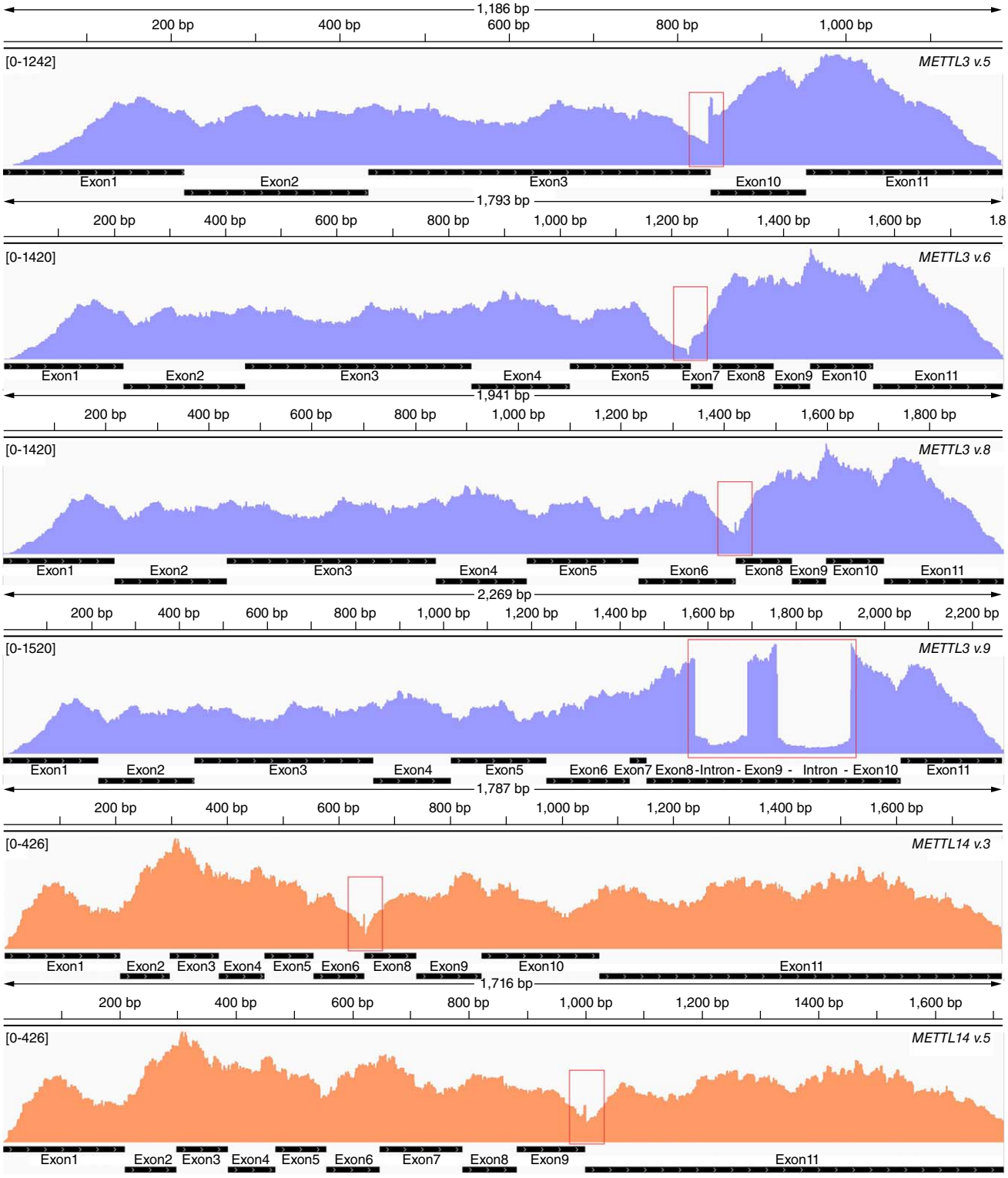


Figure 5. Ribosome profiling peaks as visualized by IGV. Red rectangles are used to highlight the region of the novel splicing event.

expected protein-coding mRNAs, *METTL3* X.1 and *METTL3* v.8, conserve all the characteristic METTL3 domains, but in a perceived position shift, due to exon skipping events (Supplementary Figure S5). AlphaFold3 predictions revealed that METTL3 X1 and v.8 have 3D structures highly similar to v.1 (Figure 6 & Supplementary Table S3), with RMSD values of 44.228 and 36.999, respectively, suggesting a strong conservation of structural features necessary for mRNA interaction and adenine methylation. Three additional putative proteins encoded by *METTL3*

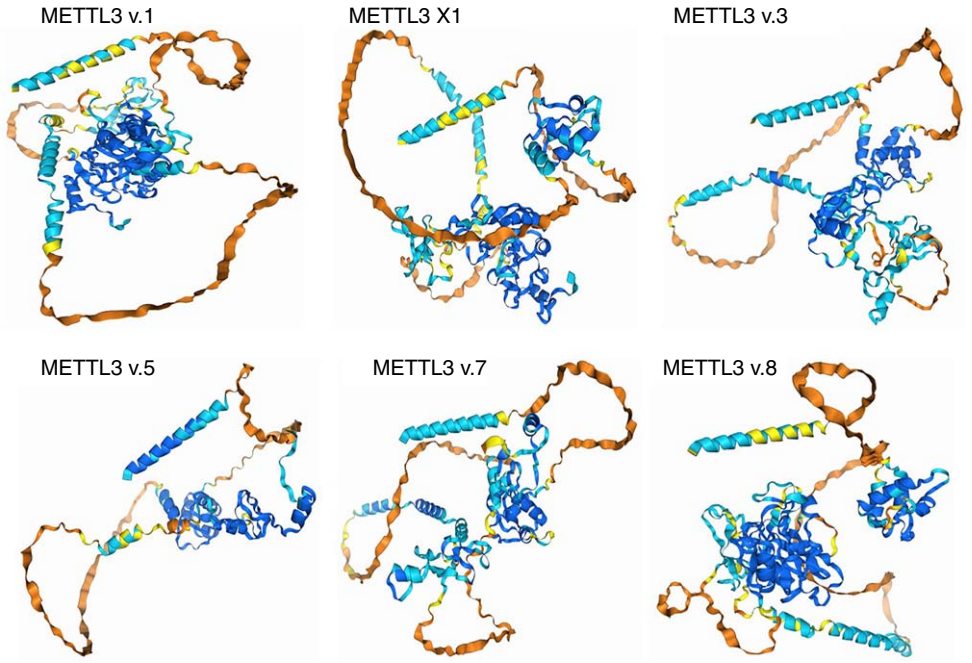


Figure 6. Protein 3D structures of the described *METTL3* isoforms obtained by AlphaFold3 software. Different colors are used to indicate the per-atom confidence on a 0–100 scale (blue: >90, cyan: >70 and <90, yellow: >50 and <70, orange: <50).

v.3, v.5 and v.7 harbor most of these functional domains. METTL3 v.3 and v.7 encompass the WTAP binding domain, the two ZFD domains and the SAM binding site but lack the DPPW domain, resulting in RMSD values of 15.449 and 23.718, respectively. Consequently, although the isoforms are expected to interact with the splicing factor WTAP and the DRACH motif found in mRNAs, the lack of the DPPW hinders the methylation of the adenine base. As for the isoform encoded by *METTL3* v.5, which showed median structural similarity (RMSD 31.807), it lacks both the ZFD domains and the DPPW region thus is expected to possess decreased methylase activity (Supplementary Figure S5). DeepFRI analysis corroborated these findings, by analyzing the GO terms of the putative protein isoforms based on their structure. For METTL3, all putative proteins are predicted to be localized into the nucleus, since they exhibit high prediction scores. Additionally, the molecular functions of the newly descripted METTL3 proteins are notably reduced or lost compared with the canonical isoform. Specifically, except for METTL3 X1 and v.8 which conserve some of the molecular functions of METTL3, including the transferase and the catalytic activity, the remaining isoforms do not possess enzymatic functions (Supplementary Table S3). In the same manner, sequence-based analysis using the DeepGOWeb server, revealed that METTL3 X1 and v.8 isoforms exhibit a similar pattern in all three GO terms as compared with the annotated METTL3 protein, whereas the prediction scores for the function of the putative products of *METTL3* v.3 and v.7 are lower. The expected isoform from *METTL3* v.5 demonstrates a notably decreased activity in terms of both molecular function and biological process, suggesting that it is mainly related to the cellular anatomical entity (Supplementary Figures S7A & S8A).

Regarding METTL14, the putative isoform encoded by v.4 conserves all six functional regions of the main isoform (Supplementary Figure S6), and structural predictions show a high degree of similarity to v.1 (RMSD 21.202). In contrast, METTL14 v.5, v.7, and v.9 exhibit significant structural deviations (RMSD values of 19.023, 8.881, and 0.701, respectively) and lack key functional motifs such as the SAM binding site, C-terminal motif region, and RGG/RG motif (Supplementary Table S4). Structure-based analysis for the METTL14 isoforms demonstrated enriched prediction scores for METTL14 v.5, v.7 and v.9 indicating their significant roles as functional enzymes (Figure 7). More specifically, all these isoforms exhibit similar scores in terms of methyltransferase and catalytic activity, acting on RNA, suggesting that METTL14 and its alternative proteins participate in the methylation machinery. Moreover, METTL14 v.5 and v.7 exhibit high prediction scores in terms of Biological Process indicating that these isoforms are involved in RNA metabolic processes, methylation and ribosome biogenesis (Supplementary Table S4). The functional profile of METTL14 v.4 exhibits significantly diminished methyltransferase activity, hence the isoform barely participates in biological processes such as the RNA

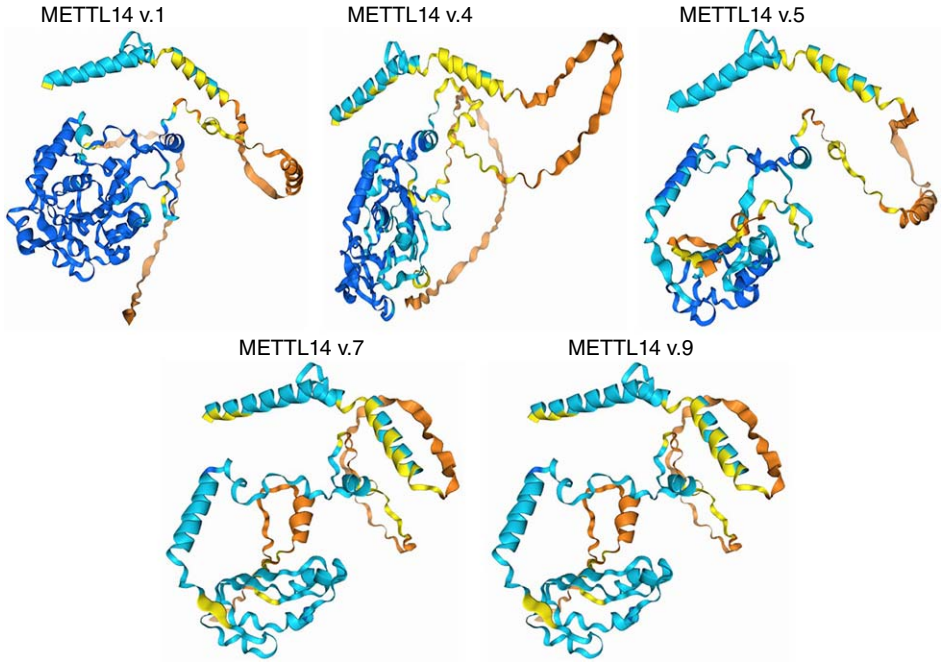


Figure 7. Protein 3D structures of the described *METTL14* isoforms obtained by AlphaFold3 software. Different colors are used to indicate the per-atom confidence on a 0–100 scale (blue: >90, cyan: >70 and <90, yellow: >50 and <70, orange: <50).

metabolic process and the RNA processing and modification (Supplementary Table S4). According to sequence-based analysis, GO terms revealed that the putative METTL14 isoforms exhibit a similar profile in all processes and molecular functions, however their enrichment scores are notably reduced compared with METTL14 (Supplementary Figures S7B & S8B).

4. Discussion

RNA splicing constitutes a key mechanism for increasing the diversity of proteins. Protein-coding genes can give rise to multiple mRNA splice variants, possibly due to different regulatory mechanisms or cellular conditions affecting transcription and mRNA processing [24]. Especially, studies support that a gene that includes 11 exons can generate an average of 5.4 alternative transcripts [25,26]. In the same manner, the purpose of the present study was to investigate the transcriptional landscape of two crucial components of the m6A methylation machinery, METTL3 and METTL14. A key novelty of our research lies in the discovery of new mRNA transcripts associated with these methyltransferases, which also contribute to the broader field of epitranscriptomics. By integrating transcriptomics data, we can generate hypotheses about the functional implications of m6A modifications and prioritize candidate transcripts for further experimental validation. Undoubtably, further proteomic-based research is needed to fully elucidate the functionality and the roles of the described mRNAs.

In the current study, we designed and performed a hybrid-seq approach by coupling the long-read nanopore sequencing and the short-read NGS to generate accurate and full-length sequencing reads that represent novel mRNAs of the major methyltransferase genes *METTL3* and *METTL14*. Our approach unveiled eight additional transcripts of *METTL3* (*METTL3* v.2-v.9) and METTL14 (*METTL14* v.2–v.9) genes. Both methyltransferases participate in the construction of m6A methyltransferase complex, which mediates mRNA stability, RNA splicing, and translation [27]. Thus, investigating the functional differences between the *METTL3* and *METTL14* alternative transcripts could uncover additional regulatory mechanisms in gene expression and mRNA metabolism. Interestingly, our *in silico* analysis strongly suggests that the described novel mRNAs are composed of both protein-coding and lncRNAs, since many transcripts are predicted to harbor PTCs and hence are highly expected to follow the pathway of NMD. The fact that half of the alternative mRNAs represent lncRNAs suggests a rather bifunctional role of *METTL3* and *METTL14* pre-mRNAs which serve as precursors of both types of RNA [28].

Since RNA-seq studies support the tissue-specific profiling of splicing events in a plethora of genes that contribute to carcinogenesis, we utilized a multiplexed nanopore sequencing approach to reveal differential expression patterns across cell lines from distinct human malignancies that were investigated [29,30]. Our analysis showed cancer-specific expression signatures and splicing events, such as novel splice junctions and intron

retentions, contributed to isoform diversity, with differential expression patterns observed across the investigated cancer types and normal cells. Of note, nanopore sequencing unveiled that *METTL3* v.9 and *METTL14* v.5 exhibited significantly higher expression levels compared with other alternative mRNAs across various human cancers and non-cancerous cells, suggesting that they may play crucial roles in cellular processes, including m^6A modification and epigenetic regulation. This finding was supported not only by nanopore sequencing data that enabled the detection and quantification of the identified full-length mRNAs but also from NGS datasets that revealed the high abundancy of the intron retention in *METTL3* v.9, and in case of *METTL14* v.5, the abundancy of the alternative splice junction between exon 9 and 11 (Figure 3). Although NGS fails to detect the full mRNA sequence of the *METTL* variants due to the short-read technology, these findings highlight the complementary nature of different sequencing technologies and underscore the necessity of comprehensive approaches in unraveling the complexity of gene expression regulation.

As for qPCR analysis, the consistency of transcript detection across diverse human cell lines suggests that these alternative transcripts are not cell type-specific and may play roles in a variety of cellular contexts. The successful amplification and detection of these transcripts in both normal and cancerous cell lines highlight their potential relevance in both conditions. On the contrary, a recent study supports the presence of 13 alternative *METTL3* mRNAs in HepG2 that exhibit differentiated expression levels between normal human liver and hepatocellular carcinoma [31]. Both *METTL3-D* (described in the study of Xu et al.) and *METTL3* v.9, which is described in our study, include the retention of the intronic regions between exons 8 and 9 and exons 9 and 10, but they differ in their fourth exon. Interestingly, both transcripts exhibit a high expression level in both studies, indicating that they are associated with human malignancies.

The detection of splicing events leading to alternative *METTL3* and *METTL14* mRNAs suggests the presence of previously unidentified protein isoforms that could possess diverse biological functions. It should be noted that despite several studies suggest ribosome's ability to recognize multiple ATG sites on mRNA, generating upstream or downstream ORFs and enabling translation of alternative in-frame and out-of-frame ORFs [32,33], our analysis was based on the annotated ATG for both *METTL3* and *METTL14*, since there is no experimental validation of alternative functional initiation sites. Our investigation of two translatome profiling studies provided substantial evidence supporting the protein-coding potential of the identified METTL transcripts. Both Poly-Ribo-Seq and TRAP-seq datasets confirmed several splicing events for both *METTL3* and *METTL14*. Importantly, Ribo-seq analysis identified peaks corresponding to these splice junctions, indicating that ribosomes are bound to these regions, implying translational activity. However, this alone may not conclusively prove that these transcripts produce functional proteins. For instance, the classification of METTL3 v.6 and v.9 as lncRNAs based on ORF query analysis raises the possibility that these transcripts may also have non-coding functions or could be translated into short, possibly non-functional peptides. The observed discrepancies might be due to alternative start codons, or the potential suppression of PTCs during translation. While the Ribo-seq data provides strong evidence for the translation potential of these alternatively spliced transcripts, further validation is necessary to confirm their protein-coding nature. These findings highlight the intricate regulatory mechanisms underlying METTL3 and METTL14 transcript diversity and their translation. The active translation of these alternatively spliced transcripts underscores their potential biological significance, warranting further investigation into their specific roles and functions in cellular processes.

One important aspect to consider is how these alternative transcripts may affect the activity or localization of the METTL3/METTL14 complex. Alternative transcripts encoding proteins with different subcellular localization signals could potentially target the methyltransferase complex to different cellular compartments, leading to spatially regulated m^6A methylation of mRNA transcripts. For METTL3, our *in-silico* analysis, incorporating AlphaFold3 structural predictions, revealed the existence of multiple putative protein isoforms resulting from alternative splicing events. Notably, certain isoforms, such as the ones derived from *METTL3* X.1 and v.8, feature all characteristic domains essential for m^6A methyltransferase activity. These findings suggest that despite the observed variations in splicing patterns, these isoforms may retain the ability to interact with mRNAs and promote methylation of adenine residues. Additionally, isoforms like *METTL3* v.3 and v.7 exhibited partial conservation of functional domains, indicating potential roles in mRNA regulation despite differences in methyltransferase activity. However, isoforms such as the one derived from METTL3 v.5 displayed notable alterations in functional domains, potentially resulting in decreased methylase activity (Figure 5). Similarly, our analysis of METTL14 isoforms identified putative protein isoforms with varying domain compositions. In detail, *METTL14* v.4 encodes a putative protein that retains all functional regions of the main isoform, while isoforms of *METTL14* v.5 and v.9 exhibited alterations in functional domains (Figure 6). Furthermore, the absence of specific

domains and the altered 3D structure of the isoforms may influence their ability to interact with other proteins or RNA substrates involved in m^{6}A modification.

To further explore the putative roles of the newly identified METTL3 and METTL14 protein isoforms, DeepFRI was utilized to annotate their functional implications based on 3D structural data. Our analysis revealed that METTL3 isoforms X1 and v.8 exhibit mild enrichment in transferase activity and catalytic activity acting on a protein related to cellular organization, biogenesis and RNA metabolism, reinforcing their role in mRNA modification and processing pathways (Supplementary Table S3). Notably, the remaining putative proteins demonstrated reduced functional predictions, correlating with their structural deviations and lack of key functional domains. For METTL14, DeepFRI highlighted conserved functions in METTL14 v.5, v.7 and v.9 isoforms More specifically, enrichment analysis revealed increased methyltransferase and catalytic activities indicating the significant role of these proteins into the methylation machinery isoform. However, although METTL14 v.4 demonstrated a preserved structural integrity, structure-based analysis unveiled reduced functional predictions (Supplementary Table S4).

It should be noted that recent studies have shown that *METTL3* knockout mESC cells express alternative mRNAs that encode protein METTL3 isoforms [31,34]. More precisely, by utilizing the CRISPR/Cas9 system to target METTL3, Poh et. al. support that alternative spliced *METTL3* transcripts can encode functional METTL3 isoforms, which possess reduced methyltransferase activities [34]. This study is in accordance with our GO bioinformatics analysis, in which the enrichment scores in terms of molecular function are lower for the putative METTL isoforms compared with the METTL proteins. The dynamic nature of transcriptional regulation is evidenced by the multitude of factors that influence gene expression, including transcription factors, epigenetic modifications and chromatin remodeling [35]. Alternative splicing enhances the complexity of the transcriptome and proteome without the need for additional genes [36]. This process allows for the generation of multiple mRNA and protein isoforms from a single gene, thereby contributing to cellular diversity and functional specialization. In the case of METTL3 and METTL14, which are integral components of the m^{6}A methylation machinery, alternative splicing can lead to the production of different protein variants with potentially distinct biological functions. Numerous studies in model organisms, such as *Drosophila melanogaster*, *Mus musculus* and *Caenorhabditis elegans*, have illustrated the critical impact of alternative splicing on developmental processes, stress responses and disease pathogenesis [37–39]. Indicatively, studies have shown that tissue-specific splicing events can regulate crucial pathways in development and differentiation, while aberrant splicing patterns are often associated with pathological conditions such as cancer and neurodegenerative diseases [40]. The regulatory complexity observed in *METTL3* and *METTL14* splicing underscores the importance of tightly controlled gene expression mechanisms and highlights the potential for therapeutic interventions targeting splicing variants.

In conclusion, our study leverages the power of transcriptomics to discover new mRNA transcripts associated with METTL3 and METTL14, highlighting the importance of comprehensive gene expression analyses in unraveling the complexities of RNA modification and its impact on gene regulation. This work not only contributes to the growing body of knowledge in epitranscriptomics but also sets the stage for future investigations into the functional roles of these newly identified transcripts.

5. Conclusion

Undoubtedly, the diversity of *METTL* alternative transcripts underscores the complexity of splicing events in the context of m^{6}A epigenetic regulation. Our findings spotlight the intricate regulation of *METTL3* and *METTL14* expression through alternative splicing, with potential implications for understanding various biological processes such as m6A epigenetic regulation in human cells. Exploring the expression patterns and regulation of the identified alternative *METTL3* and *METTL14* transcripts could provide new insights into the key factors and signaling pathways that control m^{6}A methylation dynamics. For instance, transcription factors or signaling pathways might control alternative transcript expression in response to different cellular stimuli or environmental cues. Although our study strongly supports the translating potentials of the protein-coding alternative *METTL3/14* mRNAs, the identification and structural characterization of the METTL3/14 protein isoforms is pivotal for elucidating their functional roles and implications in biological processes related to N6-adenosine-methyltransferase activity. Further functional studies are needed to elucidate the specific roles of the alternative mRNAs and their putative isoforms in gene expression regulation, cellular processes and disease pathogenesis.

Article highlights

- Nanopore sequencing unveils novel alternative *METTL3* and *METTL14* mRNAs.
- In total, eight novel *METTL3* mRNAs (*METTL3* v.2–v.9) and eight novel *METTL14* mRNAs (*METTL14* v.2–v.9) were identified.
- Expression profiling of the *METTL3/14* mRNAs was performed in human malignant and non-cancerous cells.

- *METTL3* v.9 exhibits significantly higher expression levels across the investigated human malignancies.
- *METTL14* v.5 emerged as the most abundant alternative mRNA in the tested cancer types.
- *METTL3* v.2–v.9 and *METTL14* v.2–v.9 were successfully quantified through variant-specific qPCR assays.
- Ribosome profiling supports that specific *METTL3/14* mRNAs are highly promising to encode protein isoforms.
- *In silico* structural analysis highlights the methyltransferase activity in most of the described putative novel isoforms.

Author contributions

Experiment implementation: K Athanasopoulou and PG Adamopoulos; Bioinformatics analysis: K Athanasopoulou and PG Adamopoulos. Study conception and design: PG Adamopoulos. Analysis and interpretation of results: K Athanasopoulou and PG Adamopoulos. Draft manuscript preparation: K Athanasopoulou and A Scorilas. Critical reviewing and editing: PG Adamopoulos and A Scorilas.

Financial disclosure

The authors have no financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript. This includes employment, consultancies, honoraria, stock ownership or options, expert testimony, grants or patents received or pending, or royalties.

Competing interests disclosure

The authors have no competing interests or relevant affiliations with any organization or entity with the subject matter or materials discussed in the manuscript. This includes employment, consultancies, honoraria, stock ownership or options, expert testimony, grants or patents received or pending, or royalties.

Writing disclosure

No writing assistance was utilized in the production of this manuscript.

ORCID

Konstantina Athanasopoulou
https://orcid.org/0000-0002-4071-6076
Panagiotis G. Adamopoulos
https://orcid.org/0000-0003-3939-357X
Andreas Scorilas https://orcid.org/0000-0003-2427-4949

References

Papers of special note have been highlighted as: • of interest; •• of considerable interest

1. Boo SH, Kim YK. The emerging role of rna modifications in the regulation of mrna stability. Exp Mol Med. 2020;52(3):400–408. doi:10.1038/s12276-020-0407-z
2. Xie GS, Richard HT. M(6)a mrna modifications in glioblastoma: emerging prognostic biomarkers and therapeutic targets. Cancers (Basel). 2024;16(4):727. doi:10.3390/cancers16040727
3. Liu J, Yue Y, Han D, et al. A mettl3-mettl14 complex mediates mammalian nuclear rna n6-adenosine methylation. Nat Chem Biol. 2014;10(2):93–95. doi:10.1038/nchembio.1432
4. Scholler E, Weichmann F, Treiber T, et al. Interactions, localization, and phosphorylation of the m(6)a generating mettl3-mettl14-wtap complex. RNA. 2018;24(4):499–512. doi:10.1261/rna.064063.117
5. Su X, Lu R, Qu Y, et al. Methyltransferase-like 3 mediated rna m(6) a modifications in the reproductive system: potentials for diagnosis and therapy. J Cell Mol Med. 2024;28(4):e18128. doi:10.1111/jcmm.18128
6. Su W, Che L, Liao W, et al. The rna m(6)a writer mettl3 in tumor microenvironment: emerging roles and therapeutic implications. Front Immunol. 2024;15:1335774. doi:10.3389/fimmu.2024.1335774
7. Zhang L, Xia J. N6-methyladenosine methylation of mrna in cell apoptosis. Mol Neurobiol. 2023; 61:3934–3948. doi:10.1007/s12035-023-03813-x
8. Patil DP, Chen CK, Pickering BF, et al. M(6)a rna methylation promotes xist-mediated transcriptional repression. Nature. 2016;537(7620):369–373. doi:10.1038/nature19342
9. Li J, Yang F, Wang Z, et al. Mettl16-mediated n6-methyladenosine modification of soga1 enables proper chromosome segregation and chromosomal stability in colorectal cancer. Cell Prolif. 2023;57:e13590. doi:10.1111/cpr.13590
10. Satterwhite ER, Mansfield KD. Rna methyltransferase mettl16: targets and function. Wiley Interdiscip Rev RNA. 2022;13(2):e1681. doi:10.1002/wrna.1681
11. Yan X, Liu F, Yan J, et al. Wtap-virma counteracts dsdna binding of the m(6)a writer mettl3-mettl14 complex and maintains n(6)-adenosine methylation activity. Cell Discov. 2023;9(1):100. doi:10.1038/s41421-023-00604-5
12. Kueck NA, Ovcharenko A, Hartstock K, et al. Chemoenzymatic labeling of rna to enrich, detect and identify methyltransferase-target sites. Methods Enzymol. 2021;658:161–190. doi:10.1016/bs.mie.2021.06.006
13. Fu M, Blackshear PJ. Rna-binding proteins in immune regulation: a focus on ccch zinc finger proteins. Nat Rev Immunol. 2017;17(2):130–143. doi:10.1038/nri.2016.129
• **Provides an overview of the conserved domains of METTL3-METTL14**
14. Huang J, Dong X, Gong Z, et al. Solution structure of the rna recognition domain of mettl3-mettl14 n(6)-methyladenosine methyltransferase. Protein Cell. 2019;10(4):272–284. doi:10.1007/s13238-018-0518-7
• **Provides an overview of the conserved domains of METTL3-METTL14**
15. Sledz P, Jinek M. Structural insights into the molecular mechanism of the m(6)a writer complex. Elife. 2016;5:e18434. doi:10.7554/eLife.18434
16. Xu Y, Zhang Y, Luo Y, et al. Novel insights into the mettl3-mettl14 complex in musculoskeletal diseases. Cell Death Discov. 2023;9(1):170. doi:10.1038/s41420-023-01435-9
17. Wang P, Doxtader KA, Nam Y. Structural basis for cooperative function of mettl3 and mettl14 methyltransferases. Mol Cell. 2016;63(2):306–317. doi:10.1016/j.molcel.2016.05.041

18. Li H. Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics. 2018;34(18):3094–3100. doi:10.1093/bioinformatics/bty191
19. Adamopoulos PG, Theodoropoulou MC, Scorilas A. Alternative splicing detection tool-a novel perl algorithm for sensitive detection of splicing events, based on next-generation sequencing data analysis. Ann Transl Med. 2018;6(12):244. doi:10.21037/atm.2018.06.32
20. Kulmanov M, Zhapa-Camacho F, Hoehndorf R. Deepgoweb: fast and accurate protein function prediction on the (semantic) web. Nucleic Acids Res. 2021;49(W1):W140–W146. doi:10.1093/nar/gkab373
21. Abramson J, Adler J, Dunger J, et al. Accurate structure prediction of biomolecular interactions with alphafold 3. Nature. 2024;630:493–500. doi:10.1038/s41586-024-07487-w
22. Gligorijevic V, Renfrew PD, Kosciolek T, et al. Structure-based protein function prediction using graph convolutional networks. Nat Commun. 2021;12(1):3168. doi:10.1038/s41467-021-23303-9
23. Lejeune F. Nonsense-mediated mrna decay at the crossroads of many cellular pathways. BMB Rep. 2017;50(4):175–185. doi:10.5483/BMBRep.2017.50.4.015
24. Cui H, Shi Q, Macarios CM, Schimmel P. Metabolic regulation of mRNA splicing. Trends Cell Biol. 2024;S0962-8924(24):00025–4. doi:10.1016/j.tcb.2024.02.002 (Epub ahead of print).
25. Piovesan A, Caracausi M, Antonaros F, et al. Genebase 1.1: a tool to summarize data from ncbi gene datasets and its application to an update of human gene statistics. Database (Oxford). 2016;2016. doi:10.1093/database/baw153
26. Liu Q, Fang L, Wu C. Alternative splicing and isoforms: from mechanisms to diseases. Genes (Basel). 2022;13(3):401. doi:10.3390/genes13030401
27. Wen T, Li T, Xu Y, et al. The role of m6a epigenetic modifications in tumor coding and non-coding rna processing. Cell Commun Signal. 2023;21(1):355. doi:10.1186/s12964-023-01385-w
28. Liu M, Zhang S, Zhou H, et al. The interplay between non-coding rnas and alternative splicing: from regulatory mechanism to therapeutic implications in cancer. Theranostics. 2023;13(8):2616–2631. doi:10.7150/thno.83920
29. Zheng HC, Xue H, Sun HZ, et al. The potential oncogenic effect of tissue-specific expression of jc polyoma t antigen in digestive epithelial cells. Transgenic Res. 2023;32(4):305–319. doi:10.1007/s11248-023-00352-y
30. Lee EJ, Noh SJ, Choi H, et al. Comparative rna-seq analysis revealed tissue-specific splicing variations during the generation of the pdx model. Int J Mol Sci. 2023;24(23):17001. doi:10.3390/ijms242317001
•• Provides related findings regarding alternative splicing in *METTL* genes.
31. Xu RY, Ding Z, Zhao Q, et al. An alternatively spliced variant of mettl3 mediates tumor suppression in hepatocellular carcinoma. Genes (Basel). 2022;13(4):669. doi:10.3390/genes13040669
•• The following studies provide related findings regarding alternative splicing in *METTL* genes.
32. Kochetov AV, Ahmad S, Ivanisenko V, et al. Uorfs, reinitiation and alternative translation start sites in human mrnas. FEBS Lett. 2008;582(9):1293–1297. doi:10.1016/j.febslet.2008.03.014
33. Orr MW, Mao Y, Storz G, et al. Alternative orfs and small orfs: shedding light on the dark proteome. Nucleic Acids Res. 2020;48(3):1029–1042. doi:10.1093/nar/gkz734
34. Poh HX, Mirza AH, Pickering BF, et al. Alternative splicing of mettl3 explains apparently mettl3-independent m6a modifications in mrna. PLoS Biol. 2022;20(7):e3001683. doi:10.1371/journal.pbio.3001683
35. Zheng W, Gianoulis TA, Karczewski KJ, et al. Regulatory variation within and between species. Annu Rev Genomics Hum Genet. 2011;12:327–346. doi:10.1146/annurev-genom-082908-150139
36. Du Y, Cao L, Wang S, et al. Differences in alternative splicing and their potential underlying factors between animals and plants. J Adv Res. 2023;S2090-1232(23):00354–5. doi:10.1016/j.jare.2023.11.017 (Epub ahead of print).
37. Ma Z, Sharma R, Rogers AN. Physiological consequences of nonsense-mediated decay and its role in adaptive responses. Biomedicines. 2024;12(5):1110. doi:10.3390/biomedicines12051110
38. Moroy T, Heyd F. The impact of alternative splicing *in vivo*: mouse models show the way. RNA. 2007;13(8):1155–1171. doi:10.1261/rna.554607
39. Singh P, Ahi EP. The importance of alternative splicing in adaptive evolution. Mol Ecol. 2022;31(7):1928–1938. doi:10.1111/mec.16377
40. Bonner EA, Lee SC. Therapeutic targeting of RNA splicing in cancer. Genes (Basel). 2023;14(7):1378. doi:10.3390/genes14071378

Personalized Medicine

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/ipme20

Revolutionizing personalized cancer treatment: the synergy of next-generation sequencing and CRISPR/Cas9

Muniba Mahmood, Izza Taufiq, Sana Mazhar, Faiqa Hafeez, Kausar Malik & Samia Afzal

To cite this article: Muniba Mahmood, Izza Taufiq, Sana Mazhar, Faiqa Hafeez, Kausar Malik & Samia Afzal (2024) Revolutionizing personalized cancer treatment: the synergy of next-generation sequencing and CRISPR/Cas9, Personalized Medicine, 21:3, 175-190, DOI: 10.1080/17410541.2024.2341610

To link to this article: https://doi.org/10.1080/17410541.2024.2341610

Published online: 06 May 2024.

PERSONALIZED MEDICINE
2024, VOL. 21, NO. 3, 175–190
https://doi.org/10.1080/17410541.2024.2341610

Taylor & Francis
Taylor & Francis Group

REVIEW

Revolutionizing personalized cancer treatment: the synergy of next-generation sequencing and CRISPR/Cas9

Muniba Mahmood[a], Izza Taufiq[a], Sana Mazhar[a], Faiqa Hafeez[a], Kausar Malik[a] and Samia Afzal[*,a]

[a]Centre for Excellence in Molecular Biology, University of the Punjab, Lahore, Punjab, 53700, Pakistan

ABSTRACT

In the context of cancer heterogeneity, the synergistic action of next-generation sequencing (NGS) and CRISPR/Cas9 plays a promising role in the personalized treatment of cancer. NGS enables high-throughput genomic profiling of tumors and pinpoints specific mutations that primarily lead to cancer. Oncologists use this information obtained from NGS in the form of DNA profiling or RNA analysis to tailor precision strategies based on an individual's unique molecular signature. Furthermore, the CRISPR technique enables precise editing of cancer-specific mutations, allowing targeted gene modifications. Harnessing the potential insights of NGS and CRISPR/Cas9 heralds a remarkable frontier in cancer therapeutics with unprecedented precision, effectiveness and minimal off-target effects.

ARTICLE HISTORY
Received 26 October 2023
Accepted 8 April 2024

KEYWORDS
Cancer; CRISPR/Cas9; next generation sequencing; NGS; personalized medicine

1. Background

Cancer is a clonal cell-based disease characterized by an abnormal neoplastic condition that results in tumor mass formation in the body tissues. It is the leading cause of mortality in populations worldwide owing to the dysregulation of cell cycle checkpoints and disruption of cell signaling pathways primarily associated with the cellular proliferation, metastatic invasion, evasion to apoptotic resistance, induction of angiogenesis, genomic instability, mutagenesis, abnormal telomerase activity, deregulation of the growth suppressors, destabilization of cellular energetics, and evading immune responses [1]. Various genetic mutations and epigenetic factors act differentially in different patients forming different combinations and hence are considered to be the foremost cause of disease complexity and heterogeneity of tumors [2]. As the disease progresses, cancer heterogeneity eventually increases multiple folds and makes the diagnosis and treatment highly complicated and challenging.

Conventional treatment strategies are typically based on biopsy and laboratory testing involving histological analysis and morphological examination of the tumor tissue, followed by any one, or a combination of therapies–such as chemotherapy, surgery, and radiotherapy – depending upon the type, location, and nature of the tumor [3]. The cancer treatment does not follow a 'one-size-fits-all' approach so it cannot be optimized as each case varies from patient to patient. The underlying mechanism of tumor production and progression is highly complex, as each type of tumor is caused as a result of entirely different genomic mutation and therefore some patients respond efficiently to the therapy while others remain totally unresponsive. A very low prognosis rate has been observed in patients who do not respond to the conventionally chosen therapies, due to which they develop highly metastatic and aggressive cancers leading to high risk of morbidity and relapse [4]. These challenging limitations have highlighted the significance of alternative approaches over conventional (refractory) treatment.

The convergence of next-generation Sequencing (NGS) and CRISPR/Cas9 technology represents a paradigm shift in the landscape of personalized cancer treatment. This synergy harnesses the power of high-throughput genomic analysis and precise gene editing, offering unprecedented insights into the molecular intricacies of cancer. This introduction sets the stage for exploring the transformative potential of integrating NGS and CRISPR/Cas9 in tailoring cancer therapies to individual patients, revolutionizing the field and providing new avenues for effective treatments.

Interestingly, the recent advances in high-throughput NGS and bioinformatics have made available the massive amount of significant data related to cancer genomics that has ultimately spurred the rapid development of immune therapy and targeted therapy in specifically those patients who are highly prone to metastatic invasion and aggressive cancers. A current study has shown the complete disappearance of tumor in a breast cancer patient when subjected to immunotherapy giving the high success rate, which otherwise was completely

CONTACT Samia Afzal samiaraza@live.com

unresponsive to prior chemotherapy [5]. In addition to NGS, genome editing techniques such as CRISPR/Cas9 have remarkably emerged as therapeutic tools for such complex diseases, particularly breast cancer, myelomas, non-small-cell lung carcinoma, leukemia, and glioblastoma [6]. Integrating CRISPR/Cas9 with NGS has raised the potential to identify, validate, and speed up the targeting of high-value targets leading to successful advancements in cancer treatment [7].

NGS technology has provided massive amounts of data from patients suffering from acute myeloid leukemia; later on, these data were extensively used in The Cancer Genome Atlas project designed for the analysis of solid tumors [8,9]. The data collected from this huge project led the foundation of a molecular classification system for cancer diagnosis, progression, and treatment – particularly for glioblastoma and lung cancers – as compared with the prior conventional histological classification [10]. Data obtained from different cancers have been compiled to create an NGS profile leading to the development of targeted therapy mainly involving detection of mutations in cell signaling pathways, and then using existing or highly specific novel drugs to block them. The mutations identified were either classified as 'driver or druggable' mutations if they were inevitable for cancer progression or 'passenger or nondruggable' mutations if they were not directly linked to the maintenance of cancer. This targeted approach led to the production of the constitutively active tyrosine kinase inhibitor imatinib, acting as paradigm of targeted cancer therapy and tested recently for the treatment of leukemia [11].

In the age of genomics, the advent of NGS has ushered in an era of unprecedented data acquisition, enabling us to decipher the intricate genetic tapestry woven within each cancerous lesion. Simultaneously, CRISPR/Cas9, with its precision and versatility in gene editing, has emerged as a powerful tool to dissect the genetic underpinnings of diseases, offering avenues for targeted interventions previously deemed unlikely. The amalgamation of these technologies signifies not just a convergence of methodologies but a paradigm shift in our approach to cancer therapy.

CRISPR-based genome editing creates double-stranded breaks in particular genomic sequences, followed by DNA repair mechanisms. There are two basic DNA repair mechanisms chosen on the basis of cell state and the availability of repair template: homologous recombination (specifically called homology-directed repair) and nonhomologous end joining [12]. CRIPSR/Cas9 is highly specific due to the presence of gRNAs that specifically interact with targeted sequences via the Watson–Crick base pairing rule [13]. Although the CRISPR technique has remarkable benefits and great potential, delivery of CRISPR/Cas9 editing machinery into the cells *in vivo* and avoiding off-target binding is still a key hurdle when using in therapeutics [14]. Hence, alternative high-accuracy strategies principally based on CRISPR have rapidly emerged, such as cytosine base editors, prime editing, and CRISPRon. These have surpassed conventional CRISPR-based editing as they have high efficacy and reduced off-target effects [15].

The overarching aim of this exploration is to navigate the complexities of personalized cancer treatment, spotlighting the fusion of NGS and CRISPR/Cas9 as a catalyst for groundbreaking advancements. Beyond merely decoding the genetic sequences of tumors, this integration allows us to venture into the realm of targeted therapies tailored to the unique genetic signatures of individual patients. It unravels the mystery of why certain cancers resist conventional treatments and paves the way for interventions that are not only more efficacious but also possess the potential to revolutionize the very fabric of cancer care.

2. Precision medicine perspective & cancer

Precision medicine aims to develop a specialized treatment plan precisely tailored to an individual's genetic makeup keeping in view the integrated information of a person's overall health record, lifestyle, whole genome status, and associated genetic factors [16]. It negates the one-size-fits-all approach and revolutionizes treatment strategies by designing therapy for a specific patient [17]. Precision medicine is a highly targeted approach that not only spares the patient from going through a period of detailed therapeutic procedures that have limited treatment effect on the patient but also proves to be highly cost-effective as it minimizes the excessive use of healthcare provisions and medical services [18,19].

The primary goal to personalize molecular medicine is to target highly specific mutations directly while reducing the nonspecific off-target effects [20]. NGS provides the pivotal milestone to personalized medicine by aiding the development of targeted cancer therapy. Using the NGS single-nucleotide resolution ability, tumor-specific mutations are identified through target analysis as shown in Table 1 [21]. NGS firstly analyzes a person's genetic makeup, and the mutations found therein aid in the detection of therapeutic targets that are then subjected to genome editing via CRISPR [22]. The liquid biopsy biomarkers are identified and used to monitor or track the details of tumor progression. Precision medicine mainly determines treatment strategies on the basis of liquid biopsy samples (containing circulating DNA) from patients, immunological markers, and related biologi-

Table 1. Comparison of various next-generation sequencing platforms and their applications.

Sequencing technologies	Read capacity/length (base pair)	Efficiency (%)	Advantages	Applications	Ref.
Illumina	35–300	0.1–1	Highly automatic, appropriate for high-throughput homopolymer sequencing	Epigenetic analysis, sequencing of sncRNAs and transcriptome, profiling of expressed genes	[31]
454	100–700	1–1.5	Less interference and negligible deviation in sequencing	Metagenomic studies, structural analysis of genome, advanced sequencing of transcriptome	[32]
Ion PGM	200–400	0.05–0.12	Low cost, efficient, rapid, convenient to use, undeviating coverage	*De novo* gene sequencing, SNP detection, analysis of short tandem repeats, sequencing of mitochondrial DNA	[33]
SOLiD	50–85	0.05–0.01	High accuracy, precise, suitable for sequences with high Guanine-Cytosine content	Detection of SNPs, epigenome analysis, sequencing of genome and observing transcriptome	[34]

cal aspects to measure the treatment efficacy. In addition to these, DNA profiling based on proteomes, RNA, tumor, and cell-free DNA provides a potential diagnostic option to determine the precise molecular medicine [22].

3. NGS & cancer

The emergence of NGS has significantly benefited in the targeted treatment of cancer and precision medicine by utilizing immunotherapy as shown in Figure 1. The modifications in the expression of genes of cancer aid in targeted therapy. This therapy can be used in medications that can directly combat cancer cells, with the targeted genes identified through genomic profiling in oncology. Drugs designed by immunotherapy do not attack cancer directly but instead work by stimulating the immune system. Extensive studies have been done on drug development after identifying a connection between the progression of cancer and evading of immunological defense by cancer cells. Precision medicine has become a vital component in cancer therapy due to recent developments in immunogenomics and high-throughput profiling.

3.1. Genetic profiling & target identification by NGS

A vast amount of data regarding to cancer genomics has been made accessible because of advances in NGS techniques and tools of bioinformatics. This has led to the development of novel therapies for patients dealing with aggressive cancerous tumors that cannot be treated by conventional methods. Recent research revealed that immunotherapy resulted in the complete elimination of tumor in a breast cancer patient who could not be treated by many chemotherapies or radiotherapies [23].

NGS is also known as high-throughput parallel sequencing, as described in Table 2, because it is a modern technique that can enable a researcher to gather a huge amount of genetic data in a short period of time. Sequencing of 33 types of different cancers was performed in order to identify mutations and variations present in the tumors. The data were compiled and access was provided to the public after completion of the International Cancer Genome Consortium and The Cancer Genome Atlas project [24,25]. NGS techniques and RNA sequences revealing the gene expression mechanism in cancer and the transcription factors involved were combined in a detailed database in a comprehensive project known as the Cistrome Cancer [26].

Whole-genome sequencing has been improved significantly by lowering the costs and increasing the efficiency of sequencing. It can help to identify mutations in cancer which aids in the categorization of molecular subtypes of cancer and eventually leads to development of medicines [27]. For instance, a common practice to assess a person's risk of developing breast cancer can be done by identifying BRCA1 and BRCA2 mutations through genetic testing [28]. Large and comprehensive databases with a collection of genetic variants in cancer and related data are present in the Catalogue of Somatic Mutations in Cancer [29,30]. The main purpose of these cancer databases is to provide an investigation and thorough research for the identification and validation of mutations for diagnosis and therapy of cancer.

3.2. Immunotherapy & precision medicine for treatment of cancer

A treatment strategy which is specific to each patient is known as precision medicine. This therapy plan is created after analyzing data on the patient's genetic makeup, socioeconomic background, family history, and health record. This strategy has the potential to enhance the impact of therapeutic methods by avoiding any conventional treatments that have less effect on the patient's out-

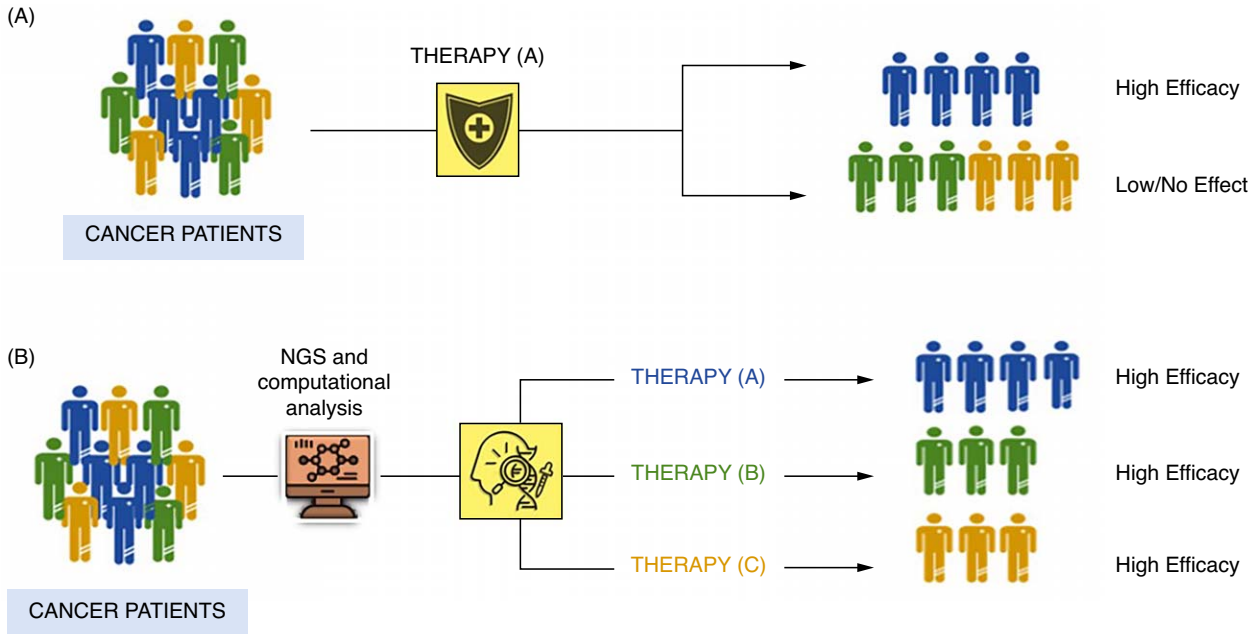


Figure 1. Comparison between the conventional versus precision therapy of cancer.
(A) Combination or sole (lone) conventional therapies including chemotherapy, surgery, and radiotherapy were applied to all cancer patients, to which some patients responded while others remained unresponsive. **(B)** Precision therapy on the basis of the individual's genetic profile obtained via NGS, immunotherapy, and CRISPR/Cas9 resulted in high treatment efficacy.
NGS: Next-generation sequencing.

Table 2. The key aspects of next-generation sequencing and CRISPR/Cas9 technologies in the context of personalized cancer treatment.

Aspect	Next-generation sequencing	CRISPR/Cas9	Ref.
Principle	High-throughput DNA sequencing for genomic analysis	Genome editing tool using RNA-guided Cas9 endonuclease	[26,27]
Advantages	Comprehensive genomic analysis	Precise gene editing at the DNA level	[15]
	Detection of genetic variations, mutations, and SNPs	Targeted modification of specific genes	[15,16]
	Enables personalized medicine through genomic profiling	Potential for therapeutic applications in genetic diseases	[12]
	Accelerates genetic and biomedical research	Versatile tool for functional genomics and drug discovery	[7]
Disadvantages	High cost associated with sequencing technology	Off-target effects may lead to unintended genetic changes	[26]
	Requires sophisticated bioinformatics for data analysis	Ethical concerns regarding germline editing in humans	[7,8]
	Data storage and management challenges	Limited efficiency in certain cell types	[68]
	Longer turnaround time for data interpretation	Delivery methods for CRISPR components can be challenging	[69]
Applications	Cancer genomics, molecular diagnostics	Gene therapy, functional genomics, disease modeling	[73]
	Infectious disease detection	Agricultural and biotechnological applications	[72,73]
	Prenatal genetic testing	Targeted cancer therapies	[72]

comes [31]. Precision medicine is also known as personalized medicine, and it aims to provide an effective cure to diseases like cancer at a reasonable cost [32].

NGS aims to assist drug design against cancer by targeted therapy. Several datasets that assess the genomic sequences, genetic variants, gene expression, and epigenetics related to cancer biology can help oncologists to develop a therapy against cancerous tumors. Conventional treatments for cancer include chemotherapy, radiotherapy, and surgery, while a more recent approach is immunotherapy guided by NGS [33]. Immunotherapy can be utilized either by itself or along with traditional cancer treatment methods to increase its effectiveness. The immune response can be enhanced and triggered against the cells by including immunotherapy in the treatment strategy [34,35].

Immunogenomics is considered a relatively novel field in the research of cancer biology. The collection of genomic profiles of immune cells and cancer cells has been made possible by the advanced technique of NGS [36] as illustrated in Figure 2. Advancements have been developed recently in single-cell RNA sequencing and chromatin immunoprecipitation sequencing technologies that play a vital role in revealing the heterogeneity of transcriptome in cancer cells. High variation levels of expression in genes related to proliferation and hypoxia were detected in glioblastoma by performing single-cell RNA sequencing [37]. The effectiveness and efficiency of

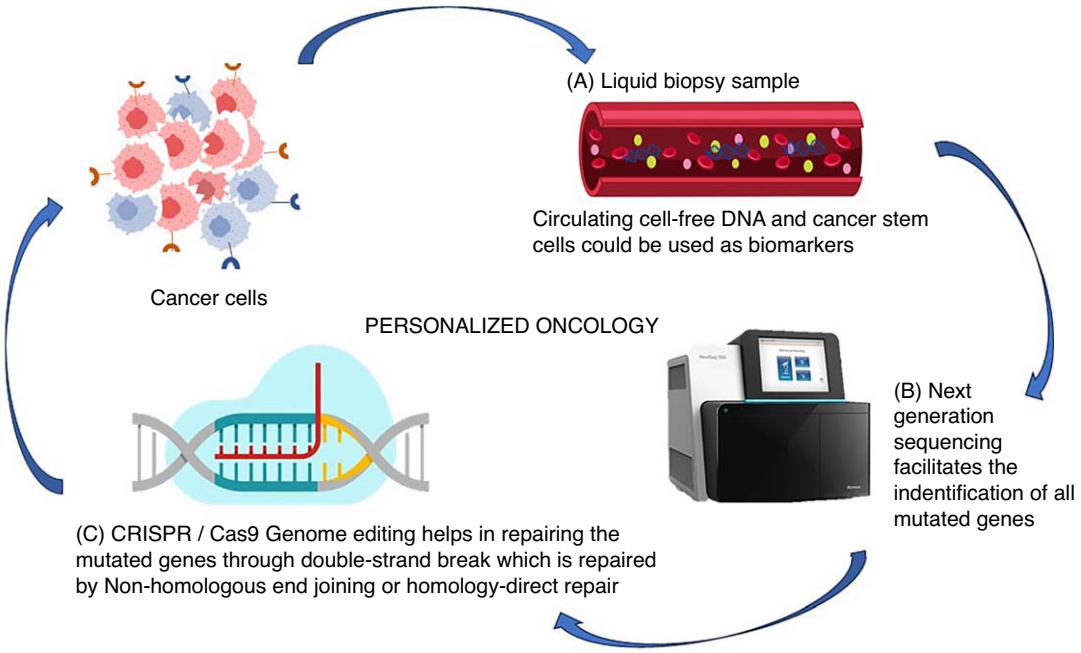


Figure 2. A brief summary of the steps involved in personalized oncology utilizing NGS and CRISPR genome editing technologies.
NGS: Next-generation sequencing.

immunotherapy can also be improved by a better understanding of cells' genetic profiles.

3.3. NGS & chimeric antigen receptor T-cell therapy

Chimeric antigen receptor (CAR) T-cell therapy is a recently developed type of adoptive T-cell therapy in which the immune response of T cells is increased by genetically altering them with CARs [38]. *Ex vivo* modification of T cells (i.e., outside the body) is performed and a new protein or CAR is expressed which targets the antigen present on the surface of cancer cells. The recognition and binding of antigen to T cell is enhanced because of the fusion between an intracellular T-cell receptor (TCR)-derived domain and an external variable fragment of monoclonal antibody, by the CAR protein [39]. These altered T cells are again introduced in the patient's body to stimulate an effective immune response.

Several clinical studies have revealed that CAR T-cell therapy is highly successful in patients suffering from leukemia and it has recently been tested for treating other types of cancers [40,41]. More than 90% individuals having acute lymphoblastic leukemia obtained remission after undergoing CAR T-cell therapy . CD19 is exclusively targeted by CAR T cells, which is expressed at significant levels only in B cells [42]. However, lack of knowledge about neoantigens or target antigens that are cancer specific is a major obstacle for utilizing CAR T-cell therapy in personalized medicine.

3.4. Vaccine development by NGS: guided immunotherapy

Foreign antigens are eliminated from the body by the attack of the immune system. Vaccines are made effective because immune cells have the ability to recognize any foreign pathogen and develop antibodies against the antigen. Mutated proteins are expressed on the surface of cancer cells called neoantigens [43]. MHC molecules present these processed mutated neoantigens which are then recognized by T cells [44]. This approach is applied in the development of vaccines by utilizing immunotherapy in which antigens secreted by cancer cells are used. Antibodies are produced against these antigens because the adaptive immunity is stimulated by the defense system [45].

Mutations in cancer cells are mapped and organized in the cancer genomic profile by employing NGS technology, which can help to predict and identify antigens from cancer cells. Recombinant proteins can be modulated to synthetically produce antigens, which can be introduced through vaccination and eventually activate the immune system [46]. Therefore, vaccines can be customized against cancer and precision medicine can be developed by utilizing the neoantigens.

4. Genome editing redefined: the CRISPR/Cas9 revolution

Due to its precision, effectiveness, and simplicity, the CRISPR/Cas9 system has proved to be a very beneficial tool for gene editing with better process timing. Cas9 in this system is an endonuclease enzyme that can introduce a double stranded cut at any site in the DNA when properly guided by a ssRNA, as shown in Figure 3. This RNA is called an sgRNA: it is approximately 20 bases in length and is complementary to the targeted DNA sequence. There is another marker present downstream of the targeted cut site called a protospacer adjacent motif; this is essential for guidance of the CRISPR complex and catalysis of cleavage at desired target site [47]. The Cas9 nuclease has two catalytic domains, HNH and RuvC, each responsible for cleaving a part of the DNA [48]. After the complementary base pairing of gRNA with target DNA, Cas9 nuclease cuts at the site adjacent to the protospacer adjacent motif sequence result in a double-stranded break.

The cell's own DNA repair mechanism repairs the introduced break. One of the ways this occurs is through nonhomologous end joining, an error-prone mechanism that can introduce deletions or insertions at the cut site causing changes in the frame shift; hence the expression of the gene via transcription or translation will be disturbed and the function of that gene will be turned off/knocked out from cell [49]. Another method is homology directed; it is a less error-prone method that requires the presence of a DNA template for repair. On the basis of the provided DNA template (*trans* DNA), highly specific insertions or DNA modifications can be made in accordance with the nicked DNA end homologous to the *trans* DNA [50]. That *trans* gene can carry a reporter, marker, or other element to control gene expression [51]. The *in vivo* expression of Cas9 nuclease can be made by delivering mRNA or DNA using an appropriate delivery system such as adeno-associated viral vectors [52].

Three components – Cas9, gRNA, and *trans*-acting RNA – form the CRISPR complex. gRNA and *trans* RNA exist in the form of a duplex. For easier and more efficient editing, this duplex can be replaced with single-stranded synthetic guided RNA [51]. This system was first discovered in bacteria as their natural defense mechanism; since then scientists have exploited this system to perform gene editing in humans [53]. There are many studies that have used CRISPR to elucidate tumor growth, heterogenicity, and etiology to come up with a therapeutic strategy, targeting only cancerous cells. For example, utilization of the CRISPR system for deletion of the *GLO1* gene in prostate cancer has helped to study disease etiology [54].

4.1. CRISPR in targeted gene therapy

Genetic information of the tumor cells provided by NGS helps us better comprehend the complexity and heterogeneity of tumor cell genomes, their etiology, and microenvironment. This information can further help us to design efficient, precise, and targeted therapy using the CRISPR/Cas9 system. Cancer is one of the leading causes of death in the world. Study of cancer has revealed the involvement of various genetic and epigenetic factors in tumor progression. Conventionally used techniques for treating cancer such as chemotherapy and even targeted molecular drugs have limitations and side effects like the effect on nontumor cells. Hence, considering the high demand for an effective and specific cancer therapy, researchers are now using CRISPR to make targeted modifications in cancerous cells [55].

CRISPR can be used in a variety of ways depending on the strategy planned and needs of the situation. The simplest approach is gene knockout, which has great potential in clinically targeting genes as in the case of HIV gene editing of receptor CCR5 enables helper T cells to better recognize the pathogen to help combat the infection [56]. Insertion of a DNA fragment can help to restore the normal sequence in the cells, as in the correction of IL2RA mutation by targeted insertion, which can treat autoimmune disease [57]. Tumors can be due to translocation of a fragment of a chromosome, so modeling this chromosomal translocation via CRISPR in an animal model can help researchers to better understand the mechanism of tumorigenesis [58]. The leading cause of many pathogenic mutation is a base/point mutation [59]. CRISPR base editing has made it possible to treat these point mutations by using adenine or cytosine deaminases in their complexes [60]. Several strategies to target cancer cells using CRISPR are explained below.

4.2. Cancer cell-specific promoters

Cancer cells have specific types of promoters on their surface that control the overexpression of genes inside the cell for tumor growth, such as MCM5, hTERT, and RAN [61]. They can be used in CRISPR for the expression of Cas enzyme or targeting of genes specifically in cancerous cells. However, there is a problem related to the less expression of these promoters but they show expression in considerable amount in healthy cells that can cause CRISPR to act efficiently in healthy cells as well. This issue can be resolved by using a combination of two promoters for the expression of a single resultant gene [62]. One promoter that can be organ specific will express the Cas9 enzyme, but it cannot function unless another promoter that will be cancer specific will express gRNA [63]. In this way CRISPR will act in a specific

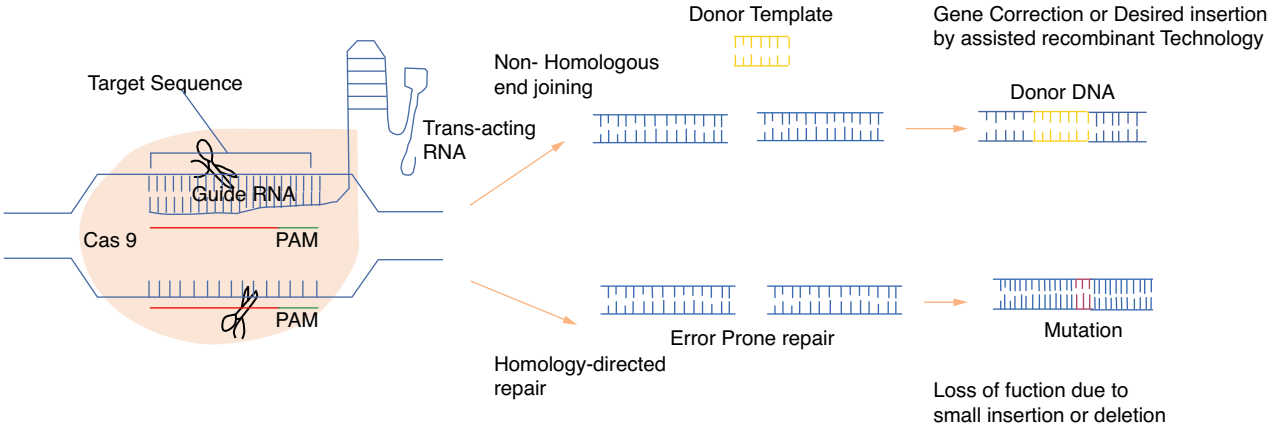


Figure 3. The mechanism of genome editing by CRISPR/Cas9 tool.

organ particularly in cancerous cells. Similarly, a second promoter-coded CRISPR system can target the *Lac1* gene in bladder cancer. This system codes for the inhibitor, thus CRISPR will activate effector genes like *p21* and *hBAX* that will cause growth halt, apoptosis, and decreased mobility in the cancer cells of bladder [64].

4.3. Telomerase & cancer: therapeutic insights

The telomerase enzyme is responsible for adding a telomere sequence at the end of DNA sequences to prevent the loss of a functional gene after each replication cycle [65]. This enzyme is expressed in highly dividing cells and, except for tumors, it is expressed in only stem cells or germline cells. CRISPR can target telomerase in cancerous cells; this involves utilization of two gene sets, one coding for gRNA and dCas9-VP64 protein and the other coding for catalytically active Cas9 attached to the DNA sequence that is recognized by telomerase and extended by telomeres in highly dividing cancer cells [66]. This extended sequence is recognized by gRNA and it then promotes the expression of nuclease Cas9 gene via dCas9-VP64. This Cas enzyme uses the same gRNA complementary to the telomere sequence to target telomeres and cause their degradation and shortening [67]. Eventually the cancer cells will lose some vital genetic information and die. This strategy has proved to be useful in cell cultures and in a mouse model with no effect in healthy living cells [68].

4.4. Cancer-specific miRNA

Different cancer cells have unique miRNAs, which can halt the expression of certain genes by binding to and degrading their mRNAs [69]. A study was performed using on 'OFF switch' design by introducing a miRNA specific sequence at one end of the Cas enzyme, RNA such that cells expressing specific miRNA would break the Cas messenger RNA. The researchers also developed an 'ON switch' design that was suitable for targeted gene editing [70]. For example, L7Ae protein in cells blocks the expression of Cas protein if the RNA of the Cas enzyme is attached to the L7Ae binding domain. As miRNA controlling the expression of L7Ae protein will block the expression of that protein, the Cas protein will be expressed easily in the cells having that specific miRNA and not in noncancerous cells [71].

4.5. CRISPR/Cas9 for oncolytic viruses

There are some viruses that integrate their genome in the host and induce consistent expression and cell proliferation. Such tumor cells express viral genes that can be a suitable target for anticancer therapy by CRISPR, as in the case of human papillomavirus gene E6 and E7, which are responsible for malignant cervical cancer [72]. The CRISPR system targeting these genes has proved to be effective by triggering apoptosis in cancer cells while healthy cells remain unaffected [73]. There are many viruses involved in different cancers in humans, such as hepatitis C and B causing liver cancer [74], Epstein–Bar virus causing pharyngeal cancer [75], and Merkel cell polyomavirus causing skin cancer [76]. Work on these cancers is still going on to look for a more effective target for CRISPR to act for a cancer cure.

4.6. Different Cas9 delivery methods & enzyme variants can limit target editing

Double-stranded break introduced by CRISPR/Cas9 allows disruption of the gene function or introduction of the *trans* gene at a targeted location. However, there is a high risk associated with the introduction of this break at the off-target site, thus disrupting the function of unintended genes. To avoid off-target deletions, insertions, or translocations, the CRISPR/Cas9 system needs to be highly specific [77]; one of the sim-

plest ways is to introduce a preformed ribonucleoprotein, Cas9/gRNA complex into the body so that the CRISPR system is ready to work and ribonucleoprotein particles degrade quickly in the body, so the time period for CRISPR and its components to stay in the body decreases so it is less likely to cause any unintended modifications but it is stable enough to offer the desired targeted editing in genome.

Other ways of dealing with the disadvantage of off-site targeting include Cas9 enzyme variants like nickase Cas9 (nCas9). It has a mutation in one of the catalytic domains, HNH or RuvC, and as a result a cut will be introduced in only one strand of DNA at a time. Utilization of two CRISPR/nCas9 complexes to target a specific site will provide greater precision as it requires both the complexes to be at the same site for a double-stranded break in the DNA [78].

4.6.1. CRISPR base editors for single base pair transitions

CRISPR/Cas9 complex for targeting specific DNA and an enzyme targeting a single strand of DNA to introduce modification in the nucleotides make up the base editor [79]. Base editing can change splice site, and may introduce point mutation or modify the sequence of amino acids. Pyrimidine can be substituted with pyrimidine base and purine can be substituted with purine bases. Enzymes that are commonly used are cytosine deaminase and adenosine deaminase. During the process, cytosine deaminase converts cytosine to uracil that is the read by the DNA replication machinery as thymine and the new nucleotide is added accordingly in nascent DNA. Similarly, adenosine deaminase converts adenine to inosine that is read by DNA machinery as guanine. Base editing works in a small specified gene window provided and guided by gRNA of CRISPR system [80]. Cytosine deaminase along with inactive Cas9 were tested in the *Escherichia coli* and human cells for off-target effects and results were quite promising with efficient targeted transitions [81]. Combination of both enzymes can also be used to have all four types of transitions at the desired site.

4.6.2. CRISPR prime editing for deletions, transversions & insertions

CRISPR prime editing can introduce any sort of base modification such as deletion, insertions, and even transversion without having to introduce a double-stranded break. It covers the limitation of base editing as it can even exchange a pyrimidine with purine and *vice versa*. In this, nCas9 is attached with reverse transcriptase enzyme, and this complex is guided to the target side by a prime gRNA [82]. This gRNA binds to the specific site of the nicked DNA that will behave as a primer and RNA attached to gRNA will encode the DNA template by using reverse transcriptase enzyme. As a result, a new nucleotide sequence will be added into the DNA: this created mismatch can be corrected by using another nCas9/CRISPR complex that will target an unedited strand of DNA. Eventually the DNA repair mechanism will repair the nick by adding nucleotides complementary to the previously edited strand. Prime editing allows deletion and substitution up to 80 and 44 base pairs respectively [83]. Thus, this modified version of CRISPR gives more a specific and precise approach for genome editing.

4.7. Regulating gene expression by CRISPR/Cas9 inactivation

Catalytically inactivated Cas9/CRISPR system can locate the specific DNA by using gRNA and can activate or suppress the expression of that gene by associating with it [84]. This system works by recruiting suppressor or enhancers to the target site. As in case of CRISPR activation, the CRISPR complex is bound to the activator like VP64 for enhanced gene expression [85]. Similarly, for suppression the CRISPR complex is bound to interfering molecules like KRAB that downregulate gene expression by interacting with its start site [86].

In addition, binding of CRISPR with DNA methyltransferase or histone modifier can permit specific epigenetic modification that opens a new way to treat cancer as there are many tumors that are related to epigenetic changes, lymphoblastic leukemia being one of the examples [87]. The role of methylation in the MCF10A of breast cancer and HCT116 of colon tumor cells has been studied. Inactivated Cas9 has been used to target tumor suppressor genes by specifically acting on their hypermethylated promoter element [88]. Hence, different molecules can be associated with CRISPR to regulate the gene expression that has a role in cancer progression.

4.8. CRISPR harboring oncolytic virus

Cas9 can produce a virus with genetic modification that can cause lysis in the cells without causing virulence [89]. Herpes simplex oncolytic virus is produced by removing virulent genes from its genome, such as ICP34.5. We can even make the virus specific to target and cause lysis in cancer cells by modification such as deletion of the *ICP6* gene (ribonucleotide reductase). Now the virus will perform oncolysis only in cells where the p16INK4A tumor suppressor gene is inactive, which is a common trait of cancer [90]. Likewise, oncolytic adenovirus produces E2F transcription factor encoded by the E1A protein to stop the cell cycle.

4.9. Therapeutic applications of CRISPR/Cas9

The prime reason for any tumor is abnormal growth of cells that can be due to inactivation of a tumor-suppressing gene or activation of oncogenes. So, the ability to manipulate genes using efficient and easy gene editing tools like CRISPR has developed a hope to design strategies to cure tumors [91]. Currently, CRISPR-based anticancer therapy is under study on cancers like breast cancer, colorectal cancer, hepatocellular carcinoma etc., targeting various genes including *TERT*, *BRAF*, and *p53* [92]. In 2016, the first ever clinical trial for tumor treatment using CRISPR/Cas9 was performed by a Chinese scientist. He knocked out the PD-1 gene for the treatment of lung cancer. Both PD-1 and PDL-1 have great importance in anticancer therapy because PDL-1 is expressed on many cancer cells. Furthermore, binding of this ligand with PD-1, expressed on our immune cells, suppress antitumor activity of immune cells and they easily bypass the defense system in the body. Since 2016, many other anticancer therapies have specifically targeted PDL-1 for diseases such as prostate cancer, liver cancer, and esophageal cancer [93].

As CRISPR can target any part of the gene for editing we can associate that capability of CRISPR with CAR T-cell therapy for successful CRISPR anticancer therapy [94]. The eye-catching ability of CRISPR has put forward three different types of immunotherapies, namely vaccine immunotherapy, checkpoint inhibition, and CAR T-cell therapy [95]. Conventionally used autologous CAR T cells have some drawbacks, such as their high cost, time-consuming nature, and difficulty to obtain the desired amount and quality of CAR T cells for patients with severe diseases. These shortcomings make the use of this therapy limited to only few patients [96]. This problem has been addressed by the CRISPR system that provides us with specifically modified allogenic or universal CAR T cells that has been derived from healthy individuals in less time with better quality [97].

CAR cells have an intracellular signaling domain that signals the activation of T cells, and their extracellular domain edited by CRISPR has the capability to specifically recognize the tumor antigen. It has been shown to be effective against multiple malignancies simultaneously, such as leukemia and lymphoma, by engineering the cells with tumor-specific receptors [98]. CAR cells also activate the release of cytokines whose level varies from person to person: an excess of the cytokines can cause neurotoxicity [99]. To deal with this drawback, T-cell receptors (TCRs) have been modified in T cells to serve the same purpose. In phase I clinical trial T cells were edited by CRISPR to remove TCRα and TCRβ, to remove their specificity, and PDL-1. Along with this phase, knockout TCRs were engineered for tumor specificity [100]. So far, this strategy has shown promising results with no neurotoxicity, as TCR does not cause release of cytokines.

CAR T-cell immunotherapy has been performed on cancers like leukemia, glioblastoma, and myeloma with successful results in animal models. Decrease in lymphocyte viability, genome instability following genome editing, and electroporation make it difficult to obtain target doses of CAR T cells using CRISPR. Moreover, the ability of CRISPR/Cas9 to act on the DNA sequences similar to the target region is another problem in manufacturing CAR T cells, as it can cause undesirable mutations in the DNA. These are the few limitations that need to be addressed for more effective control of cancer using CRISPR [101].

5. Challenges & limitations of NGS & CRISPR in precision oncology

Revolutionary technologies like NGS and CRISPR have played a significant role in precision oncology. However, in addition to their remarkable impact on science, they also have some setbacks or limitations.

5.1. Restriction of NGS in precision therapy of cancer

5.1.1. Limit of detection

In analyzing genetic mutations or changes that correspond to clinical pathology, NGS is a great tool. Yet, it has certain challenges that make it difficult to use, such as analytical sensitivity in the detection of mutation where it becomes challenging to detect mutations or a lower percentage of tumor cells because of its heterogeneity, (tumors are a form of diverse types of cell populations having different genetic makeups) [102]. Potent changes in the composition of the tumor that form of different subclonal cell types are missed over time, as this technique only analyzes the bulk type of mutations [103]. To combat this challenge, techniques like single-cell sequencing are emerging, allowing the detailed study of intratumoral heterogeneity.

5.1.2. The complexity of data analysis

This sequencing tool creates a large amount of data having all kinds of information from sources relating to genomics, epigenomics, transcriptomics, etc. The interpretation and analysis of such a massive volume of data therefore becomes laborious and challenging and it may not be accurate at all times. In NGS, errors in sequencing and systems are common. Analysis of data should be accurate, fast, and capable of detecting alteration in the genome related to cancer [104].

5.1.3. Interpretation of variants & structural variation

The major challenge that is faced in NGS is the detection of the variants that are clinically relevant. Extensive use of bioinformatics tools, analysis of various databases, and functional annotation is required to distinguish between driver and passenger mutations linked to cancer [105]. Variations such as single nucleotide and point mutation could be detected by this technique. However, large structural changes including translocation, insertion, duplication, or deletion are hard to detect. Hence, to study structural and copy number variation, considerable knowledge of bioinformatic tools is required [95].

6. Challenges of CRISPR in precision oncology

There are a number of factors that need to be considered for the best outcome of CRISPR in clinical applications. They include the delivery challenges faced by Cas9 proteins, potential threats of off-target cleavage sites, specificity of editing, multigenic diseases, and ethical considerations regarding this gene editing tool.

6.1. Off-target effect & delivery challenges

Continuous modification of the genome can generate unexpected alterations and off-target cleavage in the genome at nontargeted sites [106]. These unintended modifications on nonspecific sites can lead to unusual mutations or toxicity in cells that are being treated and reduce the specificity and selectivity of editing. Insertion and deletion at the undesired site of a gene can affect the viability of cells and may promote cancer. Thus, off-target cleavage induced by the endonuclease should be reduced while using CRISPR *in vivo* [107].

These off-target sites from gRNA endonuclease RNP were the first found *in vitro* and isolated from the genome of an animal model. The produced gRNAs are very unlikely to have off-target effects, according to the *in vivo* verification technique. However, since bioinformatic routes and various annotation algorithms are now accessible for CRISPR/Cas systems, as elaborated in Table 3, it is preferable to use them before *in vivo* and *in vitro* experiments to provide recommendations for the best gRNA design and reduce predictable off-target activities [108]. Thus, to reduce the off-target effect of CRISPR, specific or efficient ways of delivering gRNA and Cas9 proteins are to be developed.

6.2. Complex multigenic disorders & ethical consideration

Multiple pathways or genes are involved in cancer. Although CRISPR can be used to modify multiple genes or loci simultaneously, it may be necessary to use more sophisticated gene editing techniques or combinatorial methods to modify complex pathways involved in the progression of cancer utilizing CRISPR [109]. Prolonged survival and rapid growth rate distinguish cancer cells from normal cells which will indirectly help CRISPR in identifying them. There is a chance that altered cells will spontaneously undergo p53 mutation due to the ability of CRISPR/Cas9 to do so, and Cas9 can cause a p53-mediated DNA damage response.

Germline editing in precision oncology has raised ethical considerations regarding the use of this tool [110]. Proper guidelines and a check balance regarding CRISPR are necessary to ensure its responsible use. Technology advancement, ongoing research, and collaboration between different responsible bodies are required to combat these challenges in NGS and CRISPR.

The integration of NGS generates vast amounts of genomic data, posing challenges in data management, storage, and analysis. CRISPR/Cas9 raises ethical questions regarding gene editing in humans, necessitating careful consideration of its application in personalized cancer treatment. Translating genomic findings into effective personalized treatments requires overcoming hurdles in clinical implementation, including regulatory approvals and standardization. The high costs associated with NGS and CRISPR technologies may limit their accessibility for widespread use, especially in resource-constrained healthcare settings.

7. Future of precision cancer therapy by CRISPR & NGS

On the basis of biomarker analysis, precision oncology can be defined as a branch of oncology that focuses on tissue-specific, gene-directed treatments tailored according to the patient. RNA analysis, proteome, and NGS analysis are being used for DNA profiling of tumor or cell-free extract for a better understanding of immunology and to improve the ways of treating cancer [111]. The major barrier in editing techniques such as CRISPR is the presence of off-target sequences which reduce its efficacy.

Comprehensive genetic information related to specific regions of cancer or the whole genome sequence can be collected with the help of NGS [112]. The cancer-specific mutations can then be identified with the help of data analysis which serves to identify therapeutic targets. These specific genetic alterations can then be altered or modified by using CRISPR-mediated editing strategies [113].

Based on the genetic profile of different individuals, various methods of personalized treatment can be employed using techniques involving both NGS and

Table 3. An overview of how next-generation sequencing and CRISPR/Cas9 technologies contribute to various aspects of cancer treatment, from genomic profiling to drug development.

Applications	Next-generation sequencing	CRISPR/Cas9	Ref.
Genomic profiling	Identifies genetic mutations, variations and signatures in cancer cells	Enables precise identification of target genes for editing in cancer genomes	[16]
Clinical diagnostics	Facilitates molecular diagnosis, guiding treatment decisions based on genomic information	Aids in identifying specific genetic targets for personalized cancer therapies	[32]
Targeted therapies	Tailors cancer treatments based on the molecular profile of individual tumors	Offers the potential to develop targeted therapies by editing specific cancer-related genes	[68]
Drug discovery and development	Accelerates the discovery of novel cancer drugs by understanding genomic drivers	Facilitates the development of new therapeutic strategies through gene editing	[68,69]
Resistance mechanisms	Identifies mechanisms leading to treatment resistance, guiding adjustments in therapy	Offers insights into overcoming resistance by modifying genes involved in resistance pathways	[78]
Early detection and screening	Enhances early cancer detection by detecting genetic alterations at an early stage	May contribute to developing gene editing approaches for cancer prevention and early intervention	[79]
Treatment monitoring	Allows real-time monitoring of genetic changes during treatment for adaptive therapies	Holds potential for monitoring the effectiveness of gene-edited therapies in real-time	[79,80]
Limitations in hematological cancers	Challenges in detecting mutations due to the heterogeneity of hematological cancers	Application limited by the efficiency of delivery methods and off-target effects	[72]
Ethical considerations	Privacy concerns related to the handling and sharing of sensitive genomic information	Ethical considerations surrounding germline editing and unintended consequences	[78]
Cost implications	High costs associated with genomic sequencing technologies	Costs related to CRISPR-based therapies and potential affordability issues	[79]

CRISPR. NGS analysis can detect mutations that show resistance and sensitivity to particular therapies. CRISPR can then follow these findings by forming various models with the same kinds of mutations, enabling researchers to detect the effectiveness of different treatments that are specific to the person [114].

Genetic mutations that are responsible for cancer can be identified, with the help of NGS, as therapeutic targets. Against these specific sites, CRISPR employs different editing techniques to treat areas of mutations by disrupting the functions of genes responsible for the growth or progression of cancer [115]. Genetic alteration responsible for the growth of cancer and resistance to treatment can be identified through NGS. CRISPR can then be used to develop therapies against multiple pathways. Thus, effectiveness of treatment is increased.

Techniques like chromatin immunoprecipitation sequencing and single-cell RNA sequencing have improved the field, such as epigenetic heterogeneity and transcriptome analysis for the study of cancerous and immune cells at the cell level. Further studies could bring light on these sequencing techniques.

8. Opportunities in personalized medicine for therapeutic strategies

NGS allows for precise identification of genetic mutations, enabling targeted cancer therapies tailored to individual patients. The integration of CRISPR/Cas9 in drug discovery facilitates the development of novel therapies, accelerating the pace of personalized cancer treatment advancements. Genomic profiling through NGS can aid in early cancer detection, offering opportunities for preventive interventions and improved patient outcomes. CRISPR/Cas9 enables the customization of therapeutic strategies by editing specific genes, paving the way for more effective and personalized cancer treatments.

9. Conclusion

The powerful combination of NGS and CRISPR technologies opens up a wide range of revolutionizing applications for the personalized treatment of cancer. With increasing advancement and knowledge, it is crucial to acknowledge that both of these techniques are still in their early stages for the treatment of cancer and research. Many challenges have to be met before proceeding further in its implementation for humans. The gene editing ability of Cas9 has been ensured with its application in various human carcinomas, melanomas, lymphoma, etc., hence, this CRISPR/Cas9 tool can be used to edit or alter genes with the aid of NGS for identifying specific mutations which could not be found by other ways and make it feasible to understand the structural analysis of a tumor and the molecular phenomenon behind it. Yet, the economic point of view should be kept in mind, as it is one of the main hindrances in NGS. The revolutionary concept of personalized treatment for cancer is a life-changing concept to work on if technical or economic resistance can be overcome in the near future. In conclusion, while challenges exist, the synergy of NGS and CRISPR/Cas9 presents promising opportunities for revolutionizing personalized cancer treatment, ushering in a new era of precision medicine.

10. Future perspective

In the future, the combination of modern research technologies including CRISPR/Cas9 and next-generation sequencing (NGS) can have a pivotal role in transforming the therapeutic strategies for cancer. These advanced techniques show a significant potential for accurately diagnosing diseases like cancer by identifying mutations in the genes, which will help in designing personalized treatment plans. Moreover, CRISPR/Cas9 can present multiple opportunities to create precise therapies by direct modification of cancer-related genes, enhancing the efficacy of immunotherapy and combating drug resistance. However, various challenges such as unintended genetic alterations, delivery mechanisms, hurdles in the regulatory process and off-target impacts must be resolved for ensuring the efficient and safe applications of these clinical innovations. Furthermore, the interpretation and extensive analysis of genomic details and clinical information will play a significant role in revealing full potential and several benefits of personalized medicine for treatment of cancer and other diseases.

Executive summary

- This review article highlights the transformative potential of combining next-generation sequencing (NGS) with CRISPR/Cas9 technology in cancer therapy. By leveraging NGS for comprehensive tumor genome profiling and CRISPR/Cas9 for precise gene editing, this integrated approach enables personalized treatment strategies tailored to the unique genetic makeup of individual tumors. This synergy holds promise for enhancing the efficacy of cancer therapy and overcoming challenges associated with tumor heterogeneity and drug resistance, marking a significant advancement in personalized oncology.
- Precision oncology aims to tailor cancer treatment to the genetic makeup of individual tumors, enhancing therapeutic efficacy and reducing adverse effects.
- NGS enables comprehensive profiling of the cancer genome, facilitating the identification of driver mutations and molecular pathways underpinning tumor development.
- NGS allows for the identification of actionable genetic alterations in cancer, guiding targeted therapy selection and personalized treatment strategies.
- Immunotherapy harnesses the immune system to target and eliminate cancer cells, offering promising therapeutic options for patients with advanced or refractory disease.
- NGS-guided identification of tumor-specific antigens enhances the development of CAR T-cell therapies, enabling precise targeting of cancer cells while sparing normal tissues.
- NGS-guided immunotherapy facilitates the development of cancer vaccines targeting tumor-specific antigens, stimulating the immune system to recognize and eliminate cancer cells.
- CRISPR/Cas9 technology allows for precise genome editing, enabling researchers to manipulate specific genes implicated in cancer development and progression.
- CRISPR-mediated gene editing holds promise for targeted gene therapy approaches, correcting disease-causing mutations and modulating cancer-related pathways.
- Cancer cell-specific promoters enable selective targeting of therapeutic genes to cancer cells, minimizing off-target effects and enhancing treatment specificity.
- Targeting telomerase activity in cancer cells using CRISPR-based approaches represents a promising strategy for inhibiting tumor growth and proliferation.
- CRISPR-mediated modulation of cancer-specific miRNA expression offers potential therapeutic avenues for regulating oncogenic pathways and suppressing tumor progression.
- Oncolytic viruses engineered using CRISPR technology exhibit enhanced tumor specificity and oncolytic activity, providing novel therapeutic options for cancer treatment.
- Various delivery methods and Cas9 enzyme variants offer flexibility and specificity in target editing, overcoming limitations associated with off-target effects and delivery efficiency.
- CRISPR/Cas9 holds promise for a wide range of therapeutic applications in oncology, including gene therapy, cancer immunotherapy, and targeted molecular interventions.
- Despite their potential, NGS and CRISPR technologies face challenges such as data interpretation complexity, off-target effects, and delivery efficiency, necessitating ongoing research and optimization for clinical translation in precision oncology.
- In conclusion, the convergence of NGS and CRISPR/Cas9 technologies presents a transformative opportunity to advance personalized cancer treatment. Through collaborative efforts between clinicians and researchers, the integration of these innovative approaches has the potential to revolutionize cancer care, paving the way for more effective and precise therapies.

Financial disclosure

The authors have no financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript. This includes employment, consultancies, honoraria, stock ownership or options, expert testimony, grants or patents received or pending, or royalties.

Competing interests disclosure

The authors have no competing interests or relevant affiliations with any organization or entity with the subject matter or materials discussed in the manuscript. This includes employment, consultancies, honoraria, stock ownership or options, expert testimony, grants or patents received or pending, or royalties.

Writing disclosure

No writing assistance was utilized in the production of this manuscript.

ORCID

Samia Afzal https://orcid.org/0000-0002-3298-5735

References

1. Flores-Pérez JA, De La Rosa Oliva F, Argenes Y, et al. Translational research and onco-omics applications in the era of cancer personal genomics. Cham (Switzerland): Springer International Publishing; 2019. Chapter 11, Nutrition, cancer and personalized medicine; p. 157–168. https://doi.org/ 10.1007/978-3-030-24100-1_11
2. Bailey MH, Tokheim C, Porta-Pardo E, et al. Comprehensive characterization of cancer driver genes and mutations. Cell. 2018;173(2):371–385.e318. doi:10.1016/j.cell.2018.02.060

3. Urruticoechea A, Alemany R, Balart J, et al. Recent advances in cancer therapy: an overview. Curr Pharm Des. 2010;16(1):3–10. doi:10.2174/138161210789941847
4. Karantanos T, Politikos I, Boussiotis VA. Advances in the pathophysiology and treatment of relapsed/refractory Hodgkin's lymphoma with an emphasis on targeted therapies and transplantation strategies. Blood Lymphat Cancer. 2017;7:37–52. doi:10.2147/BLCTT.S105458
5. Zacharakis N, Chinnasamy H, Black M, et al. Immune recognition of somatic mutations leading to complete durable regression in metastatic breast cancer. Nat Med. 2018;24(6):724–730. doi:10.1038/s41591-018-0040-8
6. Martinez-Lage M, Puig-Serra P, Menendez P, et al. CRISPR/Cas9 for cancer therapy: hopes and challenges. Biomedicines. 2018;6(4):105. doi:10.3390/biomedicines6040105
7. Salsman J, Dellaire G. Precision genome editing in the CRISPR era. Biochem Cell Biol. 2016;95(2):187–201. doi:10.1139/bcb-2016-0137
8. Tamborero D, Gonzalez-Perez A, Perez-Llamas C, et al. Comprehensive identification of mutational cancer driver genes across 12 tumor types. Sci Rep. 2013;3(1):2650. doi:10.1038/srep02650
9. Ley TJ, Mardis ER, Ding L, et al. DNA sequencing of a cytogenetically normal acute myeloid leukaemia genome. Nature. 2008;456(7218):66–72. doi:10.1038/nature07485
10. Yang X, Lin D. [Changes of 2015 WHO histological classification of lung cancer and the clinical significance]. Zhongguo Fei Ai Za Zhi. 2016;19(6):332–336.
11. Druker BJ. Imatinib as a paradigm of targeted therapies. Adv Cancer Res. 2004;91:1–30. doi:10.1016/S0065-230X(04)91001-9
12. Cui X, Yu Y, Gupta S, et al. Autophosphorylation of DNA-dependent protein kinase regulates DNA end processing and may also alter double-strand break repair pathway choice. Mol Cell Biol. 2005;25(24):10842–10852. doi:10.1128/MCB.25.24.10842-10852.2005
13. Selvakumar SC, Preethi KA, Ross K, et al. CRISPR/Cas9 and next generation sequencing in the personalized treatment of Cancer. Mol Cancer. 2022;21(1):83. doi:10.1186/s12943-022-01565-1
14. Chen M, Mao A, Xu M, et al. CRISPR–Cas9 for cancer therapy: opportunities and challenges. Cancer Lett. 2019;447:48–55. doi:10.1016/j.canlet.2019.01.017
15. Doman JL, Raguram A, Newby GA, et al. Evaluation and minimization of Cas9-independent off-target DNA editing by cytosine base editors. Nat Biotechnol. 2020;38(5):620–628. doi:10.1038/s41587-020-0414-6
16. Morganti S, Tarantino P, Ferraro E, et al. Translational research and onco-omics applications in the era of cancer personal genomics. Cham (Switzerland): Springer International Publishing. Chapter 2, Next generation sequencing (NGS): a revolutionary technology in pharmacogenomics and personalized medicine in cancer; p. 9–30. https://doi.org/10.1007/978-3-030-24100-1_2
17. Morganti S, Tarantino P, Ferraro E, et al. Complexity of genome sequencing and reporting: next generation sequencing (NGS) technologies and implementation of precision medicine in real life. Crit Rev Oncol Hematol. 2019;133:171–182. doi:10.1016/j.critrevonc.2018.11.008
18. Brüggenjürgen B, Kornbluth L, Ferrara JV, et al. [Clinical and health economic challenges of personalized medicine]. Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz. 2012;55(5):710–714. doi:10.1007/s00103-012-1479-2
19. Kern LM, Edwards A, Kaushal R. The meaningful use of electronic health records and health care utilization. Am J Med Qual. 2016;31(4):301–307. doi:10.1177/1062860615572439
20. Chen S, Yao Y, Zhang Y, et al. CRISPR system: discovery, development and off-target detection. Cell Signal. 2020;70:109577. doi:10.1016/j.cellsig.2020.109577
21. Lancaster EM, Jablons D, Kratz JR. Applications of next-generation sequencing in neoantigen prediction and cancer vaccine development. Genet Test Mol Biomarkers. 2020;24(2):59–66. doi:10.1089/gtmb.2018.0211
22. Sgro A, Blancafort P. Epigenome engineering: new technologies for precision medicine. Nucleic Acids Res. 2020;48(22):12453–12482. doi:10.1093/nar/gkaa1000
23. Zacharakis N, Chinnasamy H, Black M, et al. Immune recognition of somatic mutations leading to complete durable regression in metastatic breast cancer. Nat Med. 2018;24(6):724–730. doi:10.1038/s41591-018-0040-8
24. Hudson Chairperson TJ, Anderson W, Aretz A, et al. International network of cancer genome projects. Nature. 2010;464(7291):993–998. doi:10.1038/nature08987
25. Tomczak K, Czerwińska P, Wiznerowicz M. The Cancer Genome Atlas (TCGA): an immeasurable source of knowledge. Contemp Oncol (Pozn). 2015;2015(1):68–77. doi:10.5114/wo.2014.47136
26. Mei S, Meyer CA, Zheng R, et al. Cistrome cancer: a web resource for integrative gene regulation modeling in cancer. Cancer Res. 2017;77(21):e19–e22. doi:10.1158/0008-5472.CAN-17-0327
27. Meldrum C, Doyle MA, Tothill RW. Next-generation sequencing for cancer diagnostics: a practical perspective. Clinical Biochem Rev. 2011;32(4):177.
28. Lynch JA, Venne V, Berse B. Genetic tests to identify risk for breast cancer. Semin Oncol Nurs. 2015. doi:10.1016/j.soncn.2015.02.007
29. Forbes SA, Beare D, Gunasekaran P, et al. COSMIC: exploring the world's knowledge of somatic mutations in human cancer. Nucleic Acids Res. 2015;43(D1):D805–D811. doi:10.1093/nar/gku1075
30. Packer BR, Yeager M, Burdett L, et al. SNP500Cancer: a public resource for sequence validation, assay development, and frequency analysis for genetic variation in candidate genes. Nucleic Acids Res. 2006;34(Suppl. 1):D617–D621. doi:10.1093/nar/gkj151
31. Brüggenjürgen B, Kornbluth L, Ferrara J, et al. Clinical and health economic challenges of personalized medicine. Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz. 2012;55:710–714. doi:10.1007/s00103-012-1479-2
32. Kern LM, Edwards A, Kaushal R, et al. The meaningful use of electronic health records and health care utilization. Am J Med Qual. 2016;31(4):301–307. doi:10.1177/1062860615572439
33. Baylin SB, Jones PA. Epigenetic determinants of cancer. Cold Spring Harb Perspect Biol. 2016;8(9):a019505. doi:10.1101/cshperspect.a019505
34. Twyman-Saint Victor C, Rech AJ, Maity A, et al. Radiation and dual checkpoint blockade activate non-redundant immune mechanisms in can-

cer. Nature. 2015;520(7547):373–377. doi:10.1038/nature14292
35. Lee Y, Auh SL, Wang Y, et al. Therapeutic effects of ablative radiation on local tumor require CD8+ T cells: changing strategies for cancer treatment. Blood. 2009;114(3):589–595. doi:10.1182/blood-2009-02-206870
36. Liu XS, Mardis ER. Applications of immunogenomics to cancer. Cell. 2017;168(4):600–612. doi:10.1016/j.cell.2017.01.014
37. Patel AP, Tirosh I, Trombetta JJ, et al. Single-cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma. Science. 2014;344(6190):1396–1401. doi:10.1126/science.1254257
38. Gomes-Silva D, Ramos CA. Cancer immunotherapy using CAR-T cells: from the research bench to the assembly line. Biotechnol J. 2018;13(2):1700097. doi:10.1002/biot.201700097
39. Gross G, Waks T, Eshhar Z. Expression of immunoglobulin-T-cell receptor chimeric molecules as functional receptors with antibody-type specificity. Proc Natl Acad Sci USA. 1989;86(24):10024–10028. doi:10.1073/pnas.86.24.10024
40. Grupp SA, Kalos M, Barrett D, et al. Chimeric antigen receptor-modified T cells for acute lymphoid leukemia. N Engl J Med. 2013;368(16):1509–1518. doi:10.1056/NEJMoa1215134
41. Kalos M, Levine BL, Porter DL, et al. T cells with chimeric antigen receptors have potent antitumor effects and can establish memory in patients with advanced leukemia. Sci Transl Med. 2011;3(95):95ra73. doi:10.1126/scitranslmed.3002842
42. Maude SL, Teachey DT, Porter DL, et al. CD19-targeted chimeric antigen receptor T-cell therapy for acute lymphoblastic leukemia. Blood. 2015;125(26):4017–4023. doi:10.1182/blood-2014-12-580068
43. Castle JC, Uduman M, Pabla S, et al. Mutation-derived neoantigens for cancer immunotherapy. Front Immunol. 2019;10:1856. doi:10.3389/fimmu.2019.01856
44. Kreiter S, Vormehr M, Van De Roemer N, et al. Mutant MHC class II epitopes drive therapeutic immune responses to cancer. Nature. 2015;520(7549):692–696. doi:10.1038/nature14426
45. Aurisicchio L, Pallocca M, Ciliberto G, et al. The perfect personalized cancer therapy: cancer vaccines against neoantigens. J Exp Clin Cancer Res. 2018;37(1):1–10. doi:10.1186/s13046-018-0751-1
46. Van Der Burg SH, Arens R, Ossendorp F, et al. Vaccines for established cancer: overcoming the challenges posed by immune evasion. Nat Rev Cancer. 2016;16(4):219–233. doi:10.1038/nrc.2016.16
47. Sternberg SH, Redding S, Jinek M, et al. DNA interrogation by the CRISPR RNA-guided endonuclease Cas9. Nature. 2014;507(7490):62–67. doi:10.1038/nature13011
48. Sgro A, Blancafort P. Epigenome engineering: new technologies for precision medicine. Nucleic Acids Res. 2020;48(22):12453–12482. doi:10.1093/nar/gkaa1000
49. Jiang F, Doudna JA. CRISPR–Cas9 structures and mechanisms. Annu Rev Biophys. 2017;46:505–529. doi:10.1146/annurev-biophys-062215-010822
50. Xia AL, He QF, Wang JC, et al. Applications and advances of CRISPR–Cas9 in cancer immunotherapy. J Med Genet. 2019;56(1):4–9. doi:10.1136/jmedgenet-2018-105422
51. He X, Tan C, Wang F, et al. Knock-in of large reporter genes in human cells via CRISPR/Cas9-induced homology-dependent and independent DNA repair. Nucleic Acids Res. 2016;44(9):e85. doi:10.1093/nar/gkw064
52. Chen M, Mao A, Xu M, et al. CRISPR–Cas9 for cancer therapy: opportunities and challenges. Cancer Lett. 2019;447:48–55. doi:10.1016/j.canlet.2019.01.017
53. Jinek M, Chylinski K, Fonfara I, et al. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. Science. 2012;337(6096):816–821. doi:10.1126/science.1225829
54. Jiang C, Meng L, Yang B, et al. Application of CRISPR/Cas9 gene editing technique in the study of cancer treatment. Clin Genet. 2020;97(1):73–88. doi:10.1111/cge.13589
55. Knott GJ, Doudna JA. CRISPR-Cas guides the future of genetic engineering. Science. 2018;361(6405):866–869. doi:10.1126/science.aat5011
56. Xu L, Wang J, Liu Y, et al. CRISPR-edited stem cells in a patient with HIV and acute lymphocytic leukemia. N Engl J Med. 2019;381(13):1240–1247. doi:10.1056/NEJMoa1817426
57. Roth TL, Puig-Saus C, Yu R, et al. Reprogramming human T cell function and specificity with non-viral genome targeting. Nature. 2018;559(7714):405–409. doi:10.1038/s41586-018-0326-5
58. Lagutina IV, Valentine V, Picchione F, et al. Modeling of the human alveolar rhabdomyosarcoma Pax3-Foxo1 chromosome translocation in mouse myoblasts using CRISPR–Cas9 nuclease. PLoS Genet. 2015;11(2):e1004951. doi:10.1371/journal.pgen.1004951
59. Landrum MJ, Lee JM, Benson M, et al. ClinVar: public archive of interpretations of clinically relevant variants. Nucleic Acids Res. 2016;44(D1):D862–S868. doi:10.1093/nar/gkv1222
60. Sato M, Takabayashi S, Akasaka E, et al. Recent advances and future perspectives of *in vivo* targeted delivery of genome-editing reagents to germ cells, embryos, and fetuses in mice. Cells. 2020;9(4):799. doi:10.3390/cells9040799
61. Rama AR, Aguilera A, Melguizo C, et al. Tissue specific promoters in colorectal cancer. Dis Markers. 2015;2015:390161. doi:10.1155/2015/390161
62. Chen C, Yue D, Lei L, et al. Promoter-operating targeted expression of gene therapy in cancer: current stage and prospect. Mol Ther Nucleic Acids. 2018;11:508–514. doi:10.1016/j.omtn.2018.04.003
63. Bowman KR, Kim JH, Lim CS. Narrowing the field: cancer-specific promoters for mitochondrially-targeted p53-BH3 fusion gene therapy in ovarian cancer. J Ovarian Res. 2019;12(1):38. doi:10.1186/s13048-019-0514-4
64. Liu Y, Zeng Y, Liu L, et al. Synthesizing AND gate genetic circuits based on CRISPR–Cas9 for identification of bladder cancer cells. Nat Commun. 2014;5(1):5393. doi:10.1038/ncomms6393
65. Akincilar SC, Unal B, Tergaonkar V. Reactivation of telomerase in cancer. Cell Mol Life Sci. 2016;73(8):1659–1670. doi:10.1007/s00018-016-2146-9
66. Mizukoshi E, Kaneko S. Telomerase-targeted cancer immunotherapy. Int J Mol Sci. 2019;20(8):1823. doi:10.3390/ijms20081823

67. Roake CM, Artandi SE. Regulation of human telomerase in homeostasis and disease. Nat Rev Mol Cell Biol. 2020;21(7):384–397. doi:10.1038/s41580-020-0234-z
68. Porika M, Tippani R, Saretzki GC. CRISPR/Cas: a new tool in the research of telomeres and telomerase as well as a novel form of cancer therapy. Int J Mol Sci. 2022;23(6):3002. doi:10.3390/ijms23063002
69. Yang J, Meng X, Pan J, et al. CRISPR/Cas9-mediated noncoding RNA editing in human cancers. RNA Biol. 2018;15(1):35–43. doi:10.1080/15476286.2017.1391443
70. Ali Syeda Z, Langden SSS, Munkhzul C, et al. Regulatory mechanism of microRNA expression in cancer. Int J Mol Sci. 2020;21(5):1723. doi:10.3390/ijms21051723
71. Hirosawa M, Fujita Y, Parr CJC, et al. Cell-type-specific genome editing with a microRNA-responsive CRISPR–Cas9 switch. Nucleic Acids Res. 2017;45(13):e118. doi:10.1093/nar/gkx309
72. Zhen S, Li X. Oncogenic human papillomavirus: application of CRISPR/Cas9 therapeutic strategies for cervical cancer. Cell Physiol Biochem. 2017;44(6):2455–2466. doi:10.1159/000486168
73. Li C, Guo L, Liu G, et al. Reprogrammed CRISPR-Cas13a targeting the HPV16/18 E6 gene inhibits proliferation and induces apoptosis in E6-transformed keratinocytes. Exp Ther Med. 2020;19(6):3856–3860. doi:10.3892/etm.2020.8631
74. Liu X, Hao R, Chen S, et al. Inhibition of hepatitis B virus by the CRISPR/Cas9 system via targeting the conserved regions of the viral genome. J. Gen. Virol. 2015;96(8):2252–2261. doi:10.1099/vir.0.000159
75. Yuen KS, Chan CP, Wong NM, et al. CRISPR/Cas9-mediated genome editing of Epstein–Barr virus in human cells. J Gen Virol. 2015;96(Pt 3):626–636. doi:10.1099/jgv.0.000012
76. Gavvovidis I, Leisegang M, Willimsky G, et al. Targeting Merkel cell carcinoma by engineered T cells specific to T-antigens of Merkel cell polyomavirus. Clin Cancer Res. 2018;24(15):3644–3655. doi:10.1158/1078-0432.CCR-17-2661
77. Kosicki M, Tomberg K, Bradley A. Repair of double-strand breaks induced by CRISPR–Cas9 leads to large deletions and complex rearrangements. Nat Biotechnol. 2018;36(8):765–771. doi:10.1038/nbt.4192
78. Ran FA, Hsu PD, Lin CY, et al. Double nicking by RNA-guided CRISPR Cas9 for enhanced genome editing specificity. Cell. 2013;154(6):1380–1389. doi:10.1016/j.cell.2013.08.021
79. Kuscu C, Parlak M, Tufan T, et al. CRISPR-STOP: gene silencing through base-editing-induced nonsense mutations. Nat Methods. 2017;14(7):710–712. doi:10.1038/nmeth.4327
80. Gaudelli NM, Komor AC, Rees HA, et al. Programmable base editing of AoT to GoC in genomic DNA without DNA cleavage. Nature. 2017;551(7681):464–471. doi:10.1038/nature24644
81. Doman JL, Raguram A, Newby GA. Evaluation and minimization of Cas9-independent off-target DNA editing by cytosine base editors. Nat Biotechnol. 2020;38(5):620–628. doi:10.1038/s41587-020-0414-6
82. Anzalone AV, Randolph PB, Davis JR, et al. Search-and-replace genome editing without double-strand breaks or donor DNA. Nature. 2019;576(7785):149–157. doi:10.1038/s41586-019-1711-4
83. Newby GA, Liu DR. *In vivo* somatic cell base editing and prime editing. Mol Ther. 2021;29(11):3107–3124. doi:10.1016/j.ymthe.2021.09.002
84. Wang Q, Zhuang P, Huang H, et al. Mouse γ-synuclein promoter-mediated gene expression and editing in mammalian retinal ganglion cells. J Neurosci. 2020;40(20):3896–3914. doi:10.1523/JNEUROSCI.0102-20.2020
85. Maeder ML, Linder SJ, Cascio VM, et al. CRISPR RNA-guided activation of endogenous human genes. Nat Methods. 2013;10(10):977–979. doi:10.1038/nmeth.2598
86. Gilbert LA, Horlbeck MA, Adamson B, et al. Genome-scale CRISPR-mediated control of gene repression and activation. Cell. 2014;159(3):647–661. doi:10.1016/j.cell.2014.09.029
87. Wang Q, Xiong F, Wu G, et al. Gene body methylation in cancer: molecular mechanisms and clinical applications. Clin Epigenetics. 2022;14(1):154. doi:10.1186/s13148-022-01382-9
88. Choudhury SR, Cui Y, Lubecka K, et al. CRISPR-dCas9 mediated TET1 targeting for selective DNA demethylation at BRCA1 promoter. Oncotarget. 2016;7(29):46545–46556. doi:10.18632/oncotarget.10234
89. Choi AH, O'leary MP, Fong Y, et al. From benchtop to bedside: a review of oncolytic virotherapy. Biomedicines. 2016;4(3):18. doi:10.3390/biomedicines4030018
90. Aghi M, Visted T, Depinho RA, et al. Oncolytic herpes virus with defective ICP6 specifically replicates in quiescent cells with homozygous genetic mutations in p16. Oncogene. 2008;27(30):4249–4254. doi:10.1038/onc.2008.53
91. Martínez-Jiménez F, Muiños F, Sentís I, et al. A compendium of mutational cancer driver genes. Nat Rev Cancer. 2020;20(10):555–572. doi:10.1038/s41568-020-0290-x
92. Topalian SL, Taube JM, Pardoll DM. Neoadjuvant checkpoint blockade for cancer immunotherapy. Science. 2020;367(6477):eaax0182. doi:10.1126/science.aax0182
93. Xu X, Liu C, Wang Y, et al. Nanotechnology-based delivery of CRISPR/Cas9 for cancer treatment. Adv Drug Deliv Rev. 2021;176:113891. doi:10.1016/j.addr.2021.113891
94. Sekar D, Lakshmanan G, Biruntha M. Implications of CRISPR/Cas9 system in Hypertension and its related diseases. J Hum Hypertens. 2021;35(7):642–644. doi:10.1038/s41371-020-0378-5
95. Mukherjee S. Genomics-guided immunotherapy for precision medicine in cancer. Cancer Biother Radiopharm. 2019;34(8):487–497. doi:10.1089/cbr.2018.2758
96. Miliotou AN, Papadopoulou LC. CAR T-cell therapy: a new era in cancer immunotherapy. Curr Pharm Biotechnol. 2018;19(1):5–18. doi:10.2174/1389201019666180418095526
97. Mollanoori H, Shahraki H, Rahmati Y, et al. CRISPR/Cas9 and CAR-T cell, collaboration of two revolutionary technologies in cancer immunotherapy, an instruction for successful cancer treatment. Hum Immunol. 2018;79(12):876–882. doi:10.1016/j.humimm.2018.09.007
98. Labanieh L, Majzner RG, Mackall CL. Programming CAR-T cells to kill cancer. Nat Biomed Eng. 2018;2(6):377–391. doi:10.1038/s41551-018-0235-9

99. Gauthier J, Turtle CJ. Insights into cytokine release syndrome and neurotoxicity after CD19-specific CAR-T cell therapy. Curr Res Transl Med. 2018;66(2):50–52. doi:10.1016/j.retram.2018.03.003
100. Stadtmauer EA, Fraietta JA. CRISPR-engineered T cells in patients with refractory cancer. Science. 2020;367(6481):eaba7365. doi:10.1126/science.aba7365
101. Salas-Mckee J, Kong W, Gladney WL, et al. CRISPR/Cas9-based genome editing in the era of CAR T cell immunotherapy. Hum Vaccin Immunother. 2019;15(5):1126–1132. doi:10.1080/21645515.2019.1571893
102. Zhang T-H, Wu NC, Sun R. A benchmark study on error-correction by read-pairing and tag-clustering in amplicon-based deep sequencing. BMC Genomics. 2016;17(1):108. doi:10.1186/s12864-016-2388-9
103. Mandelker D, Schmidt RJ, Ankala A, et al. Navigating highly homologous genes in a molecular diagnostic setting: a resource for clinical next-generation sequencing. Genet Med. 2016;18(12):1282–1289. doi:10.1038/gim.2016.58
104. Kamps R, Brandão RD, Bosch BJVD, et al. Next-generation sequencing in oncology: genetic diagnosis, risk prediction and cancer classification. Int J Mol Sci. 2017;18(2):308. doi:10.3390/ijms18020308
105. Gu W, Miller S, Chiu CY. Clinical metagenomic next-generation sequencing for pathogen detection. Ann Rev Path. 2019;14(1):319–338. doi:10.1146/annurev-pathmechdis-012418-012751
106. Cheng AW, Wang H, Yang H, et al. Multiplexed activation of endogenous genes by CRISPR-on, an RNA-guided transcriptional activator system. Cell Res. 2013;23(10):1163–1171. doi:10.1038/cr.2013.122
107. Khalaf K, Janowicz K, Dyszkiewicz-Konwińska M, et al. CRISPR/Cas9 in cancer immunotherapy: animal models and human clinical trials. Genes (Basel). 2020;11(8):921. doi:10.3390/genes11080921
108. Johnson TM. Perspective on precision medicine in oncology. Pharmacotherapy. 2017;37(9):988–989. doi:10.1002/phar.1975
109. Allmer J, Yousef M. miRNomics: microRNA biology and computational analysis Totowa (NJ): Humana; 2022. https://doi.org/10.1007/978-1-62703-748-8
110. Niemiec E, Howard HC. Ethical issues related to research on genome editing in human embryos. Comput Struct Biotechnol J. 2020;18:887–896. doi:10.1016/j.csbj.2020.03.014
111. Sekar D, Thirugnanasambantham K, Hairul Islam VI, et al. Sequencing approaches in cancer treatment. Cell Prolif. 2014;47(5):391–395. doi:10.1111/cpr.12124
112. Jamal-Hanjani M, Wilson GA, Mcgranahan N, et al. Tracking the evolution of non-small-cell lung cancer. N Engl J Med. 2017;376(22):2109–2121. doi:10.1056/NEJMoa1616288
113. Medaglia C, Giladi A, Stoler-Barak L, et al. Spatial reconstruction of immune niches by combining photoactivatable reporters and scRNA-seq. Science. 2017;358(6370):1622–1626. doi:10.1126/science.aao4277
114. Zhang H, Qin C, An C, et al. Application of the CRISPR/Cas9-based gene editing technique in basic research, diagnosis, and therapy of cancer. Mol Cancer. 2021;20(1):126. doi:10.1186/s12943-021-01431-6
115. Fierro J, Dipasquale J, Perez J, et al. Dual-sgRNA CRISPR/Cas9 knockout of PD-L1 in human U87 glioblastoma tumor cells inhibits proliferation, invasion, and tumor-associated macrophage polarization. Sci Rep. 2022;12(1):2417. doi:10.1038/s41598-022-06430-1

Master the microbiome:

microbiome standards, long-read sequencing, metagenome assembly and profiling

Microbiome standards are imperative for profiling and analyzing microbial communities. In many cases, a lack of proper controls or comparison to microbiome reference materials means that important and high-impact conclusions cannot be reproduced or reliably compared to similar datasets. In this webinar, scientists from Zymo Research and PacBio will discuss the different types of microbiome standards, how they are used, how to select the appropriate control and examples of how they have been used in long-read sequencing to benchmark microbial profiling and metagenome assembly.

Panelists

Kris Locken

Research Scientist, Zymo Research

Kris is a molecular microbiologist focused on developing and creating educational resources on microbiome standards and controls at Zymo Research (CA, USA).

Prior to joining the microbiome team at Zymo Research in 2019, he was the Quality Control Manager at two different medical device manufacturing facilities. He enjoys talking about microbiology and how cool microbes are.

Jeremy Wilkinson

Senior Staff Specialist, Microbial Genomics, PacBio

Jeremy is the Segment Lead for Microbial Genomics at PacBio (CA, USA). Jeremy came to PacBio around 2 years ago from the Harvard T.H. Chan School of Public Health (MA, USA), where he was the Director of the Microbiome Analysis Core and a Research Scientist in the Department of Biostatistics in the Huttenhower Lab for 4 years.

Watch now >>>

BioTechniques
The International Journal of Life Science Methods

Editorial Department

Digital Editor
Beatrice Bowlby
beatrice.bowlby@tandf.co.uk

Business Development and Support

Commercial Director
Sarah Mayes
sarah.mayes@tandf.co.uk

@MyBioTechniques
@biotechniques
BioTechniques

www.biotechniques.com