

Genomic Sequencing in Clinical Oncology

Aashna K. Soni

BASIS Independent Fremont, 39706 Mission Blvd, Fremont, CA 94539, USA; soniaashnakaur@gmail.com

Mentor: Gauravjit Singh

ABSTRACT: Next-generation sequencing (NGS) technologies can simultaneously sequence millions of small DNA or RNA fragments. They mark a paradigm shift over the preceding first-generation sequencing methodologies, notably Sanger sequencing. With its low cost and high throughput, NGS holds value for many different clinical applications. This review aims to discuss the application of NGS in a clinical oncology setting—namely, how sequencing is used to detect an individual's unique mutations and inform targeted therapy options. This review will focus specifically on how Foundation Medicine, a leading company in cancer biomarker detection, conducts its testing and interprets the results obtained from sequencing. Additionally, this review will explore the underlying mechanism of NGS by illustrating the Illumina sequencing workflow, one of the most commonly used NGS workflows today that is also used in Foundation Medicine's products. Finally, the review will look ahead to newer innovations and improvements that will enhance the ability of genomic sequencing to deliver tangible benefits to cancer patients.

KEYWORDS: Translational Medical Sciences; Disease Treatment and Therapies; Genomic Sequencing; Illumina Sequencing; Genomic sequencing in cancer applications.

■ Introduction

Cancer is a disease in which cells grow uncontrollably.¹ For many years, conventional chemotherapy was the standard treatment for most cancers. Chemotherapy works by targeting cells at different stages of the cell cycle.² Since cancer cells divide much more rapidly than normal cells due to the specific mutations they carry, chemotherapy is often highly successful at killing cancerous cells. However, because chemotherapy cannot distinguish between healthy and cancerous cells, normal cells are also damaged in the process, which can lead to serious side effects.² For example, because hair follicle cells divide rapidly, chemotherapy often damages these cells, resulting in significant hair loss.³

As a result, a new class of treatments, collectively known as targeted therapies, has emerged to target an individual's unique mutations by looking for characteristic biomarkers through a tumor's DNA sequence. Whereas chemotherapy kills cancer cells in the body, targeted therapy prevents cancer cells from replicating by targeting these specific biomarkers. Mechanisms by which targeted therapies work include stimulating the immune system to kill cancer cells, interrupting growth-stimulating signals from reaching cancer cells, interfering with signals that trigger angiogenesis, or causing cancer cells to undergo apoptosis.⁴ These drugs significantly improve over standard chemotherapy, demonstrating the benefit of using an individual tumor's DNA sequence to target cancer cells.

Most targeted therapies fall into two categories: monoclonal antibodies and small-molecule drugs. Monoclonal antibodies are designed to function as "substitute antibodies,"⁵ binding to antigens on cancer cells. They can serve many functions, including flagging cancer cells for destruction by the immune system, blocking proteins that cancer cells need to divide and

proliferate, or inhibiting immune checkpoints.⁵ An example of a monoclonal antibody is trastuzumab, which is used to treat human epidermal growth factor receptor 2 (HER2) positive breast cancer. This molecule attaches to HER2 on cancer cells, which prevents HER2 from sending signals that allow cancer cells to continue growing.⁶ On the other hand, small-molecule drugs have small molecular weights and can enter cancer cell membranes to target intracellular molecules.⁷ An example of a small-molecule drug is sorafenib, a tyrosine kinase inhibitor.⁸

Many targeted therapies today rely on NGS to determine a tumor's DNA sequence. For example, Foundation Medicine conducts NGS for solid tumors to suggest FDA-approved therapies. Though NGS has shown great success in oncology applications, some downsides exist. For example, the short read lengths generated from a DNA sample make it difficult to identify larger structural variations. As a result, an emerging area of interest is single-molecule sequencing, which allows for longer read lengths to be produced. Thus, single-molecule sequencing represents an area in which to explore new types of complex mutations involved in cancer, such as large insertions, deletions, inversions, duplications, and translocations. Using short-read technologies, structural variations are determined indirectly by aligning multiple reads to a reference genome.⁹ In some cases, certain types of mutations, such as copy number variations, may not even be detected from short-read data.⁹ Long read coverage, on the other hand, can allow for entire structural variations to be covered in a single read in many instances, increasing sensitivity to these types of mutations.⁹ Additionally, long-read sequencing performs particularly well in highly repetitive regions of DNA and hard-to-access regions of the genome, such as telomeres and centromeres.⁹

Most reviews focus on precision medicine in cancer focus on a specific aspect of the process, such as sequencing techniques, the development of targeted therapies, different types of biopsies in the clinic, or the genetic complexity and functional relevance of different mutation types. This review is different in that it aims to combine all of these areas to coherently describe the process of genomic sequencing and targeted therapies in a clinical context, as well as to point to newer techniques that can improve the current state of cancer care.

■ Results and Discussion

The First Generation of DNA Sequencing: Sanger Sequencing:

Frederick Sanger's chain-termination sequencing technique, developed in 1977,¹⁰ marked a significant achievement in the field of sequencing technology. An illustration of the process of Sanger sequencing is shown in Figure 1. The technique uses modified nucleotides called dideoxynucleotides triphosphates (ddNTPs), whose sugar molecules "lack a hydroxyl group on the 3' carbon,"¹⁰ preventing extension of the DNA chain. The given DNA sample is first denatured to separate the two strands.¹¹ ddNTPs, which are labeled with a different dye color for each nitrogenous base,¹¹ and dNTPs, are put together into the reaction, though the concentration of ddNTPs is less, along with DNA polymerase and the desired DNA sample to be sequenced.¹⁰ A primer is annealed to the DNA template strand, allowing DNA polymerase to extend the chain of nucleotides.¹⁰ Whenever a ddNTP is incorporated instead of a dNTP, the chain is terminated due to the lack of a hydroxyl group on the 3' carbon of the nucleotide.¹⁰ After many cycles, millions, or even billions,¹¹ of DNA fragments of different lengths are generated, guaranteeing that a ddNTP would have been incorporated at each location in the template DNA.¹⁰ These fragments undergo capillary electrophoresis, which separates the fragments based on size, allowing the terminating nucleotides on the fragments to be read in order from smallest to largest fragment.¹⁰ The results are shown on a chromatogram, in which a peak represents a nucleotide.¹⁰ By detecting a ddNTP and its characteristic dye from the end of each DNA fragment, the instrument can determine the sequence of the DNA sample nucleotide by nucleotide.¹⁰

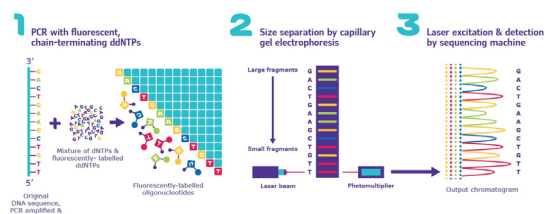


Figure 1: Sanger sequencing workflow.¹¹

Though Sanger sequencing can produce high-quality reads with low error rates, it is inefficient, expensive, and time-consuming.¹² These pitfalls opened the door for newer sequencing methodologies, which collectively fall under the category of next-generation sequencing. These sequencing technologies are described in the next section.

The Second Generation of DNA Sequencing:

The second generation of DNA sequencing represents a significant shift over preceding first-generation sequencing technologies in that it produces short reads in a massively parallel fashion.¹² Sanger sequencing was able to produce high-quality reads with low error rates.¹² However, this process is costly and inefficient.¹² On the other hand, in NGS, multiple DNA fragments are sequenced simultaneously and later assembled to generate the complete DNA sequence, making it much faster and cheaper than Sanger sequencing.¹²

There are many different types of NGS. Examples include Roche/454 sequencing, which utilizes a pyrosequencing approach, and Ion torrent sequencing, which works by detecting hydrogen ions released as DNA polymerase incorporates nucleotides.¹² However, this review will focus specifically on the Illumina NGS approach, as it is one of the most widely used workflows today and is utilized by many companies, including Foundation Medicine. The general process of Illumina sequencing consists of the following steps: library preparation, in which adapters are ligated to DNA fragments; cluster generation, in which bridge amplification is used to produce multiple copies of the DNA fragment; sequencing by synthesis, in which fluorescent signals emitted from the flow cell after each nucleotide addition are detected to determine the DNA sequence; and data analysis, in which the millions of reads generated are aligned to a reference genome and variants are determined.¹³ The first three steps of the process are shown in Figure 2. First, the given DNA sample is randomly fragmented, and adapters are added to the ends of each fragment.¹⁴ These adapters allow the fragments to hybridize to the flow cell in the next step.¹⁴ Then, through bridge amplification, in which double-stranded DNA "bridges" are formed as DNA fragments bend over and bind to vacant flow cell oligos, DNA polymerase synthesizes new strands, generating many clusters of DNA fragments.¹³ In the sequencing stage, the fluorescently-labeled bases are called through the incorporation of complementary nucleotides.¹⁴ As the flow cell is imaged, the "emission wavelength and intensity,"¹³ which are produced from each cluster, determine the base call. Then, the reversible terminator on the nucleotide is cleaved,¹⁴ allowing for the incorporation of the next base. This process is repeated many times to generate a read of the desired read length.¹³ This procedure is known as sequencing by synthesis, as the template DNA strand is read through the incorporation of complementary nucleotides by DNA polymerase.¹⁴ Finally, the millions of reads generated are aligned to the reference genome to synthesize the full sequence, and variants are identified.¹³

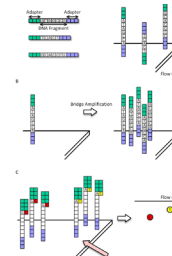


Figure 2: Illumina sequencing workflow.¹⁵

NGS in Oncology:

NGS technology has shown tremendous value in clinical cancer applications due to its ability to identify targeted therapies for patients. As the quality, efficiency, and cost of NGS have improved, more and more patients are being treated with targeted therapy and immunotherapy. The following sections describe the process of biomarker testing in a clinical context, from the biopsy to the interpretation of results.

Biopsy:

The process of cancer care begins with a biopsy, in which cells or tissues are removed from a specific location in the patient's body to be analyzed by the pathologist.¹⁶ A biopsy is necessary to confirm a cancer diagnosis, and many different types of biopsies can be administered. For this review, the two main types of biopsies that will be discussed are tissue biopsies and liquid biopsies.

Tissue biopsies are the standard for biopsies in cancer patients owing to the different types of data that can be analyzed from a tissue specimen.¹⁷ Tissue biopsies allow for a histologic analysis of the specimen and the detection and analysis of important "non-DNA-based alterations."¹⁷ Frequently, the tissue is preserved in formalin-fixed paraffin-embedded (FFPE) format. This preparation is ideal for immunohistochemistry, as "cell structures and proteins are well preserved."¹⁸ The drawback of using FFPE tissue is that the DNA and RNA in these samples are often degraded and chemically damaged during the fixation process due to chemical crosslinking between nucleic acids and proteins.¹⁹ Nevertheless, many companies, such as Foundation Medicine, use FFPE-preserved tissue in their genomic analysis process. On the other hand, fresh frozen tissue biopsies are much better for analyzing DNA or RNA, but they are difficult to store for long periods.¹⁸

The second main type of biopsy is a liquid biopsy. Liquid biopsies are performed when a tissue biopsy is infeasible or when a biopsy might pose a significant danger to the patient.¹⁷ In a liquid biopsy, cell-free DNA (cfDNA), or small DNA fragments located in "the noncellular component of the blood,"¹⁷ is isolated from bodily fluids to analyze the amount and sequence of circulating tumor DNA (ctDNA) present. Many different types of bodily fluids can be used for liquid biopsies, such as urine, cerebrospinal fluid, saliva, and blood.¹⁷ Blood is the most commonly used fluid in a liquid biopsy.¹⁷ The benefit of a liquid biopsy is that it is a minimally-invasive procedure, as compared to a tissue biopsy, which could be painful and dangerous.¹⁷ Furthermore, studies have shown that analysis of cfDNA is better able to capture the "molecular heterogeneity associated with resistance,"¹⁷ since there are often many different tumor subpopulations within a single individual, and each tumor subclone carries unique mutations. The unique mutations carried by different tumor populations in a metastatic cancer can be better captured with a blood sample that contains ctDNA released throughout the body.¹⁷ However, liquid biopsies may not always yield optimal results for patients if the concentration of ctDNA in the blood is too low, as current technologies are unable to detect ctDNA below a certain concentration.¹⁷

Analysis of the Biopsy:

After a tissue or blood sample is collected, it is sent to a pathologist, who performs different tests to provide a final diagnosis and compiles information about the gross and microscopic description of the specimen in a pathology report.²⁰ After the pathology report is issued, the medical oncologist may send the biopsy for NGS to drive a targeted treatment strategy. The results of the sequencing procedure are compiled in a biomarker testing report. Foundation Medicine's biomarker testing approach is described below.

Foundation Medicine Genomic Sequencing:

Foundation Medicine performs a type of NGS called comprehensive genomic profiling (CGP) on solid tumors. Foundation Medicine's tissue-based and blood-based tests analyze 324 genes. The combined DNA and RNA test (FoundationOne® Heme) analyzes 406 DNA genes and 265 RNA genes to better understand gene fusions and rearrangements.^{21,22} In addition to looking for mutations in a wide range of genes, Foundation Medicine analyzes the clinically significant genomic signatures of microsatellite instability (MSI) and tumor mutational burden (TMB). The specific sequencing technique used by Foundation Medicine is hybridization-based capture technology,²³⁻²⁵ which sequences a subset of genomic regions that are isolated through "hybridization to target-specific biotinylated probes."²⁶ Hybridization-based capture technology allows a large number of genes to be sequenced at a high mutation resolution to detect and understand new and known variants.²⁶

Foundation Medicine Genomic Signatures Analyzed:

Foundation Medicine analyzes two main genomic signatures in DNA samples: MSI and TMB.²¹

A microsatellite is a "short segment of DNA, usually, one to six or more base pairs in length, repeated multiple times in succession at a particular genomic location."²⁷ MSI occurs when there are differences in the number of microsatellite repeats, making the region of nucleotides unstable.²⁸ MSI occurs due to deficient DNA mismatch repair (dMMR). Mismatch repair is responsible for correcting such errors during nucleotide incorporation. MSI could indicate that an individual has Lynch syndrome, which could put them at high risk for colorectal cancer.^{29, 30}

TMB measures the number of somatic mutations "per coding area of a tumor genome."³¹ It has been shown that high TMB (TMB-H) generally occurs in cancers linked to high mutagen exposure, an example being the connection between smoking and the presence of non-small cell lung cancer (NSCLC).³¹

MSI and TMB are important biomarkers that predict a patient's response to immunotherapy treatment. One specific type of immunotherapy used for MSI-H or TMB-H cancers is immune checkpoint inhibitors (ICI). Immune checkpoints are pathways that ensure that the immune system does not attack healthy cells in the body.³² However, cancer cells utilize these pathways to their advantage to render immune cells inactive.³² A common mechanism through which immune checkpoints function is through "ligand-receptor interactions."³² In this process, a ligand on the cancer cell binds to a receptor on the T-cell, inactivating the T-cell and preventing it from attacking

the cancer cell.³² The process of ICI takes advantage of these interactions.³² In ICI treatment, monoclonal antibodies are delivered to target “negative regulators of T-cell function,”³³ or proteins that are part of the ligand-receptor interactions. Examples of such negative regulators include PD-L1 (present on tumor cells), PD-1 (present on T-cells), and CTLA4 (present on T-cells).³³ The monoclonal antibody binds to one of these proteins, preventing it from binding to the corresponding protein on either a cancerous cell or a T-cell. This allows T-cells to remain active and find and destroy cancer cells. Figure 3 illustrates the mechanism of action of immune checkpoint inhibitors, specifically targeting PD-1 or PD-L1.³⁴

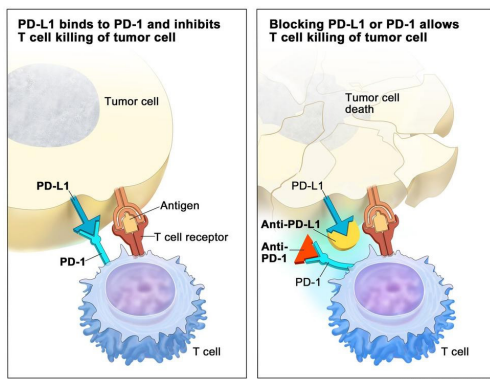


Figure 3: ICI mechanism of action.³⁴

Driver Genes: Background:

Driver genes, or genes that, when mutated, contribute to cancer development, are grouped into two main categories: proto-oncogenes and tumor suppressor genes.³⁵ Proto-oncogenes code for proteins that trigger cell division.³⁵ When mutated, proto-oncogenes become oncogenes, which code for proteins that excessively stimulate cell division.³⁶ One mechanism through which a proto-oncogene can be converted to an oncogene is when mutations within a proto-oncogene result in an overactive protein that stimulates cell division being produced in the normal amount.³⁶ Other mechanisms include multiple copies of the proto-oncogene being created and mutations within a control region, such as promoter, both of which lead to an abnormally large quantity of the normal growth-stimulating protein.³⁶ Examples of clinically important proto-oncogenes are *BRAF*, *KRAS*, and *MYC*.³⁵

On the other hand, tumor suppressor genes produce proteins that inhibit cell division.³⁶ Mutations to tumor suppressor genes destroy safeguards designed to inhibit cell division, resulting in cell proliferation.³⁶ Examples of tumor suppressor genes include *TP53*, *PTEN*, and *CDKN2A*.³⁵ Furthermore, a specific group of tumor suppressor genes functions in DNA repair.³⁶ An example of a repair mechanism that, when malfunctioning, contributes to cancer development is mismatch repair (MMR), which is primarily involved in correcting base-base errors and small indels.³⁷ Mutations to DNA repair genes increase the likelihood of cells developing other driver mutations, which further contribute to cancer development.³⁶

Foundation Medicine Genomic Alteration Classes Analyzed:

Genomic alterations that occur in driver genes play an important role in cancer development. The main genomic alteration classes analyzed by Foundation Medicine are base substitutions, insertions and deletions (indels), copy number alterations (CNAs), and select genomic rearrangements.²³⁻²⁵

Single base substitutions (SBSs) occur when one nucleotide in the DNA sequence is replaced with another.³⁸ Base substitutions can be either missense, which results in an incorrect amino acid, nonsense, which results in a premature protein, or silent, which does not affect the amino acid produced.³⁸ According to a review published in 2014, the number of SBSs present in an individual can vary greatly, both among different cancer types and within cancer types.³⁹ Cancers, like lung cancer, whose presence is linked to “chronic mutagen exposure,”³⁹ were found to carry a large number of SBSs.³⁹ SBS mutations that are quite prominent in cancer genomes are “C→T transitions at CG:CG sites and substitutions at C:G base-pairs in the context of YC:GR dinucleotides,”³⁹ where “Y” represents a pyrimidine and “R” represents a purine, and a colon separates dinucleotides on opposite strands. For example, SBS signatures found in lung cancer include C:G → A:T and T:A → G:C transversions, whose putative causes are tobacco smoke and arsenic exposure, respectively.³⁹

Indels occur when nucleotides are added to or deleted from the DNA sequence.⁴⁰ Indel mutations are usually confined to less than 1 kb of nucleotides.⁴⁰ Indels lead to a shift in the reading frame if the number of bases added or deleted is not a multiple of three, and this generally results in a nonfunctional protein.⁴¹ These types of indels are called non-3n indels. On the other hand, 3n indels, in which the number of bases added or deleted is a multiple of three, will cause amino acid changes but will not have as drastic effects as non-3n indels.⁴² According to a study published in 2010, since indels can significantly affect gene function, such as through the downstream effects of frameshift mutations, indels can serve as a type of driver mutation in cancer development.⁴² This was demonstrated by the abundance of indels in the COSMIC database analyzed in the 2010 study.⁴² Additionally, this study found that frameshift mutations, or non-3n indels, are much more prevalent in tumor suppressor genes than in oncogenes,⁴² whereas 3n indels are much more common in oncogenes, as shown by the fact that there were 7.9 times more 3n indels in oncogenes than in tumor suppressor genes when looking at genes with greater than or equal to 100 mutations.⁴² Overall, the study found that though “one- or two-bp indels and non-3n indels are dominant in both genome and coding sequences,”⁴² this can vary considerably based on the specific niche location of indels in the genome.⁴²

CNAs are a type of structural change in genomic material in which regions of the DNA sequence are deleted or amplified. CNAs can affect areas ranging from just a few kilobases up to entire chromosomes.⁴³ Somatic copy number alterations (SCNAs), or those acquired during an individual’s life and not passed on to future generations, play a pivotal role in tumorigenesis.⁴³ SCNAs allow tumor cells to increase the expression of certain genes and decrease the expression of

others, thus likely conferring a certain advantage to these cells.⁴³ The “first comprehensive pan-cancer analysis of SCNAs,”⁴⁴ published in 2013, revealed that large amplifications or deletions are commonly present in ovarian carcinomas, and the number of SCNAs varies drastically based on the tumor type.⁴⁴

Finally, rearrangements are large-scale changes to chromosome structure that include duplications, deletions, insertions, inversions, and translocations.⁴⁵ One important result of many genomic rearrangements is gene fusion. Gene fusions are “hybrid genes formed when two previously independent genes become juxtaposed.”⁴⁶ Fusions can result from both structural and non-structural rearrangements, and they are a common source of driver mutations in many cancers.⁴⁶ The mechanisms through which fusion genes contribute to tumorigenesis include causing one of the fusion partners (such as a tumor suppressor gene) to lose its function, coding for an abnormal protein that has “oncogenic functionality,”⁴⁶ or deregulating one of the fusion partners (such as a proto-oncogene).⁴⁶ The frequency of gene fusions is highly variable among different cancers.⁴⁶ One important example of gene fusion occurs in lung cancer. In patients with *ROS1*-positive lung cancer, a gene fusion occurs between the *ROS1* gene and a portion of another gene, the most common one being the *CD74* gene.⁴⁷ This fusion results in an activation of the *ROS1* gene that leads to cell proliferation.⁴⁷

Interpretation of Data:

Information found through NGS can predict which types of treatment have the highest likelihood of working for a given patient based on the driver mutations identified. For example, common driver mutations in lung cancers occur in the genes *EGFR* and *KRAS*.⁴⁸ *EGFR* is a gene that codes for the *EGFR* receptor protein involved in triggering cell proliferation and survival upon a ligand binding to it.⁴⁹ *KRAS* codes for the K-Ras protein which is involved in the RAS/MAPK signaling pathway, and it is involved in transmitting signals from the cell's environment to the nucleus to trigger cell division or differentiation.⁵⁰ Based on a patient's unique mutation, the Foundation Medicine test will recommend a targeted therapy. For example, currently, patients with the *KRAS* G12C mutation will still undergo first-line systemic therapy.⁵¹ However, if the patient does not respond well to this standard treatment, they may be prescribed the *KRAS* inhibitor, sotorasib, an FDA-approved small molecule.⁵² Sotorasib works by binding to *KRAS* G12C and inhibiting “*KRAS* oncogenic signaling.”⁵²

Genomic signatures can also guide treatment. MSI-H or TMB-H tumors would likely respond well to immunotherapy treatment, such as pembrolizumab, an immune checkpoint inhibitor. Pembrolizumab is an FDA-approved anti-PD-1 drug to treat patients with MSI-H or TMB-H cancers under specific circumstances (such as when prior treatment has not been successful), where TMB-H is defined as greater than or equal to 10 mutations/megabase.⁵⁴

If no FDA-approved therapies are available for the patient, they may be encouraged to enroll in a clinical trial to access new drugs still in development. Clinical trial enrollment has the added benefit of helping researchers gather data on the

safety and efficacy of new drugs, promoting the development of new treatment options for a broad group of patients.

Despite the significant amounts of valuable data generated by NGS, medical oncologists still look at many other factors when considering which drugs to prescribe to a particular patient, including cancer type, stage, age, and overall health. For example, an aggressive drug would likely not be used to treat an elderly patient with late-stage cancer, as the side effects could be unmanageable. Nevertheless, data collected from NGS are pivotal in identifying targeted therapies that are more likely to work than conventional chemotherapy.

Advantages of Foundation Medicine's Approach:

Because Foundation Medicine accommodates both tissue and blood samples through its FoundationOne® CDx and FoundationOne® Liquid CDx, even patients with tumors in difficult-to-access locations can benefit from the large-scale genomic analysis. Additionally, the massively parallel nature of the test allows for simultaneously testing hundreds of genes to detect many different types of biomarkers.

Proven Benefits of NGS: Non-Small Cell Lung Cancer:

A patient suspected of having lung cancer will be sent for a biopsy. The type of biopsy can vary based on the location of the tumor and the patient's specific conditions. Examples of biopsy procedures include bronchoscopy, mediastinoscopy, and transthoracic needle aspiration.⁵¹ Tissue biopsies are the standard for biopsies.⁵¹ However, liquid biopsies can be performed in cases when an invasive tissue biopsy would be unsafe for a patient based on their medical conditions.⁵¹

Let us assume that a patient is diagnosed with NSCLC. A patient with advanced or metastatic disease will have their histologic subtype established.⁵¹ Based on these results, the patient will be diagnosed with an NSCLC subtype, such as adenocarcinoma or squamous cell carcinoma.⁵¹ If the patient's histologic subtype is determined to be adenocarcinoma, then a tumor sample will be sent for biomarker testing to look for mutations in genes such as *EGFR* and *ALK*, as well as genes such as *KRAS*, *ROS1*, and *BRAF*.⁶⁰ In addition to testing for these genes, the patient's PD-L1 status will also be tested to determine the potential for immunotherapy treatment.⁶⁰ If the patient has an *EGFR* exon 19 deletion or *L858R* mutation and the “*EGFR* mutation [is] discovered prior to first-line systemic therapy,”⁶⁰ the preferred drug for the patient will be osimertinib,⁵¹ a third-generation *EGFR* tyrosine kinase inhibitor (EGFR-TKI).⁵⁵

Protein tyrosine kinases are proteins that play an important role in cell-signaling pathways that are involved in controlling “cellular differentiation and proliferation.”⁵⁶ Osimertinib works by inhibiting mutant *EGFR*, the protein product of activating mutations in *EGFR*.⁵⁵ Osimertinib differs from earlier-generation EGFR-TKIs in that in addition to targeting exon 19 deletions and exon 21 L858R mutations, it also targets the EGFR T790M mutation. This mutation often causes tumors to develop resistance to earlier EGFR-directed therapies.⁵⁷ Osimertinib has an overall response rate (ORR) of 80% based on the FLAURA clinical trial, performing significantly better than standard EGFR-TKIs.⁵⁸ However, tumors can still acquire resistance to osimertinib treatment through mechanisms such

as “loss of the T790M mutation during osimertinib therapy.”⁵⁵ If this change is accompanied by other resistance mechanisms, such as *MET* amplification and *KRAS* mutation, even first-generation EGFR-TKIs may be ineffective.⁵⁵ Based on the symptoms (or lack thereof) that the patient experiences, they may progress to a new therapy, such as surgery or stereotactic ablative radiotherapy (SABR), or may even continue osimertinib.⁵¹ Overall side effects of osimertinib include diarrhea, constipation, rash, itching, and vomiting.⁵⁹ Currently, other drugs are being studied in clinical trials to target tumors that have developed specific resistance mechanisms to osimertinib.⁵⁸

Future of Genomic Sequencing in Cancer:

Though NGS marks a significant improvement over conventional chemotherapy approaches for treating patients, it still has certain disadvantages. For example, NGS generates very short read lengths, which makes it difficult to assemble the reads to the reference genome and accurately detect structural variations (SVs).⁶⁰ Additionally, the amplification-based procedures involved often obscure important mutations that are present at a lower frequency.⁶⁰ Single-molecule sequencing (SMS) has the potential to address these gaps to make the process of genomic sequencing in cancer more fruitful. In the following section, the benefits and pitfalls of SMS are described.

Single-molecule Sequencing:

SMS, or third-generation sequencing (TGS), is a newer sequencing approach, currently spearheaded by Pacific Biosciences (PacBio) and Oxford Nanopore Technologies (ONT), that allows for much longer read lengths to be generated compared to NGS workflows.⁶⁰

Figure 4 compares NGS with two types of TGS: PacBio SMRT sequencing and ONT sequencing. The main difference between the two types of sequencing methods is the size of the read lengths. While the insert, or DNA fragment sandwiched between two adapters, is less than 300 bp in NGS,⁶⁰ PacBio can sequence tens of kilobases of DNA.⁶¹ And, Nanopore sequencing has an even longer read length capacity, sequencing as low as 20 bases all the way up to millions of bases.⁶² In PacBio SMRT sequencing, a circular DNA template is generated (as opposed to linear templates in NGS) and then sequenced as DNA polymerase incorporates nucleotides.⁶⁰ The DNA polymerase sequencing by synthesis method is a similarity between PacBio SMRT sequencing and NGS. In ONT sequencing, DNA fragments are passed through nanopores situated in an electrically resistant membrane in which an electrical current is run.⁶³ As the DNA fragment passes through the nanopore, electrical current disruptions occur, which are measured to determine the base call.⁶³

The advantage of SMS approaches is that they can sequence long fragments of DNA directly, which differs from current NGS approaches that rely on amplification techniques such as bridge amplification.⁶⁰ Additionally, they are substantially better at identifying SVs compared to conventional NGS approaches due to the long reads produced.⁶⁰ Currently, there is difficulty in identifying SVs using NGS because of the short read lengths.⁶⁴ In fact, current short-read approaches detect

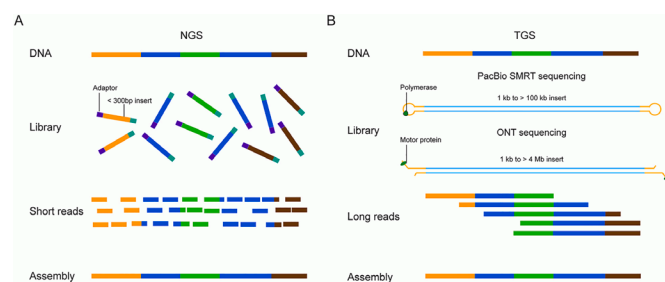


Figure 1: Comparison of NGS and TGS.⁶⁰

within the range of 10% to 70% of SVs and have false positive rates ranging up to 89%.⁶⁴ Short-read approaches are also not designed to capture the full complexity of SVs.⁶⁴ On the other hand, SMS, by utilizing longer reads, can better identify larger, more complex SVs in the given DNA sample. Overall, SMS can also detect mutations that occur at low frequencies in the DNA sample.⁶⁰

SVs, including large insertions, deletions, inversions, duplications, and translocations, represent an important class of mutations in cancer.⁶⁵ For example, some serious small cell lung cancer cases that are high-grade are driven by somatically acquired SVs.⁶⁶ There are many mechanisms through which SVs can contribute to tumorigenesis. They can alter the copy number of genes, obstruct the crucial function of tumor-suppressor genes, form fusion genes, or pair one gene's coding sequence with another's regulatory machinery.⁶⁵ In fact, new types of complex SVs are being discovered,⁶⁷ and these variants could help develop new anti-cancer drugs.⁶⁸ Sakamoto et al. found a unique class of mutations, cancerous local copy-number lesions (CLCLs), using SMS.⁶⁷ CLCLs are composed of “complex combinations of copy-number changes (duplications), inversions, and deletions,”⁶⁷ and they were found within important cancer genes “such as the *STK11*, *NF1*, *SMARCA4*, and *PTEN* genes.”⁶⁷ These types of mutations were difficult to identify and characterize solely based on short-read sequencing,⁶⁷ suggesting the need for long-read sequencing approaches to identify novel mutation types that could have clinical relevance as drug targets. Additionally, this study confirmed that long-read sequencing approaches alone can still identify small-scale mutations, such as point mutations, which are changes in single nucleotide pairs.⁶⁷

Currently, the main disadvantage of SMS is the high error rate.⁶⁰ It has been shown that for PacBio SMRT sequencing and ONT sequencing, most errors can be attributed to “false insertions or deletions.”⁶⁰ In the future, it is possible that NGS will be combined with SMS for a “hybrid sequencing strategy.”⁶⁰ This approach would allow for better quantification of fusion genes and mRNA transcript isoforms and reduce the error rate.⁶⁰ Furthermore, to successfully incorporate long-read sequencing into the clinic, the format of the tissue biopsy collected will need to change.⁶⁸ Currently, the downside of using FFPE samples is that the nucleic acids are not well preserved and experience chemical damage and fragmentation.⁶⁸ Due to this limitation, fresh frozen tissue, which does a considerably better job at preserving nucleic acid structure, is needed for long-read sequencing.⁶⁸

Therefore, with certain modifications to the process of genomic sequencing in a clinical setting, a wider variety of SVs can be determined, which could both serve as an important way to detect new mutation targets as well as tailor treatment to an individual's specific mutations.

■ Conclusion

Clinical oncology is continuously evolving as new technologies are developed and improved upon. The progress made in this field within the last few decades is commendable, and an increasing number of people are receiving personalized cancer treatment driven by genomic sequencing. Current areas of interest include incorporating newer techniques, such as SMS, to improve the detection of SVs through their ability to generate long reads. Thus, new technologies continue to expose the underlying genomic complexity of cancer that continues to humble experts in the field. However, no technology is perfect: each has its advantages and disadvantages. SMS suffers from higher error rates than NGS technologies. The future of clinical oncology likely lies in a combination of SMS and NGS, along with the development of bioinformatics pipelines that process long-read data, to leverage the unique benefits of each approach to deliver different kinds of actionable insights.

■ Acknowledgments

I want to thank radiation oncologist, Dr. Gauravjit Singh for guiding me throughout the process of writing this literature review and providing feedback on my work.

■ References

1. *What is Cancer?* – NCI. <https://www.cancer.gov/about-cancer/understanding/what-is-cancer> (accessed 2022-06-30).
2. *How Does Chemo Work?* | *Types of Chemotherapy*. <https://www.cancer.org/treatment/treatments-and-side-effects/treatment-types/chemotherapy/how-chemotherapy-drugs-work.html> (accessed 2022-07-05).
3. *Hair Loss*. <https://www.breastcancer.org/treatment-side-effects/hair-loss> (accessed 2022-07-10).
4. *Targeted Therapy for Cancer* – NCI. <https://www.cancer.gov/about-cancer/treatment/types/targeted-therapies> (accessed 2022-06-28).
5. *Monoclonal antibody drugs for cancer: How they work* – Mayo Clinic. <https://www.mayoclinic.org/diseases-conditions/cancer/in-depth/monoclonal-antibody/art-20047808> (accessed 2022-07-09).
6. Greenblatt, K.; Khaddour, K. Trastuzumab. In *StatPearls*; National Library of Medicine; StatPearls Publishing, 2022.
7. *Small molecules*. <https://www.astrazeneca.com/r-d/next-generation-therapeutics/small-molecule.html> (accessed 2022-07-23).
8. *Sorafenib: MedlinePlus Drug Information*. <https://medlineplus.gov/druginfo/meds/a607051.html#:~:text=Sorafenib%20is%20in%20a%20class,the%20spread%20of%20cancer%20cells> (accessed 2022-07-24).
9. De Coster, W.; Van Broeckhoven, C. Newest Methods for Detecting Structural Variations. *Trends in Biotechnology* 2019, 37 (9), 973–982.
10. *DNA sequencing (article)* | *Biotechnology* | *Khan Academy*. <https://www.khanacademy.org/science/ap-biology/gene-expression-and-regulation/biotechnology/a/dna-sequencing> (accessed 2022-06-14).
11. *Sanger Sequencing Steps & Method*. <https://www.sigmaaldrich.com/US/en/technical-documents/protocol/genomics/sequencing/sanger-sequencing> (accessed 2022-06-14).
12. Kchouk, M.; Gibrat, J. F.; Elloumi, M. Generations of Sequencing Technologies: From First to next Generation. *Biology and Medicine* 2017, 09 (03), 4–5.
13. Next-Generation Sequencing (NGS) | Explore the technology. *An Introduction to Next-Generation Sequencing Technology*. Illumina, Inc., 2017. https://www.illumina.com/content/dam/illumina-marketing/documents/products/illumina_sequencing_introduction.pdf (accessed 06-15-22).
14. *NGS Workflow Steps* | *Illumina sequencing working workflow*. <https://www.illumina.com/science/technology/next-generation-sequencing/beginners/ngs-workflow.html#:~:text=In%20the%20Illumina%20sequencing%20workflow,a%20process%20known%20as%20multiplexing> (accessed 2022-06-16).
15. Heo, Y. Improving Quality of High-Throughput Sequencing Reads. Ph.D. Dissertation, University of Illinois at Urbana-Champaign, Urbana-Champaign, 2015. <https://www.ideals.illinois.edu/items/89344> (accessed 2022-07-04).
16. *Definition of biopsy* – *NCI Dictionary of Cancer Terms* – NCI. <https://www.cancer.gov/publications/dictionaries/cancer-terms/def/biopsy> (accessed 2022-07-02).
17. Corcoran, R. B.; Chabner, B. A. Application of Cell-Free DNA Analysis to Cancer Treatment. *New England Journal of Medicine* 2018, 379 (18), 1754–1765.
18. *The Pros and Cons of FFPE vs Frozen Tissue Samples*. <https://www.geneticistinc.com/blog/the-pros-and-cons-of-ffpe-vs-frozen-tissue-samples> (accessed 2022-07-04).
19. Kresse, S. H.; Namlos, H. M.; Lorenz, S.; Berner, J.-M.; Myklebost, O.; Bjerkehagen, B.; Meza-Zepeda, L. A. Evaluation of Commercial DNA and RNA Extraction Methods for High-Throughput Sequencing of FFPE Samples. *PLOS ONE* 2018, 13 (5).
20. *What information is included in a pathology report?* <https://www.cancer.org/treatment/understanding-your-diagnosis/tests/testing-biopsy-and-cytology-specimens-for-cancer/whats-in-pathology-report.html> (accessed 2022-07-04).
21. *Compare our Tests* | *Foundation Medicine*. <https://www.foundationmedicine.com/portfolio> (accessed 2022-07-08).
22. FoundationOne® Heme | Foundation Medicine. *FoundationOne Heme Overview*. FoundationMedicine®, 2021. https://assets.ctfassets.net/w98cd481qyp0/4mnCQaQuXydMHziIjsnukz/45a431ee91fd81e8090c00f9e8a1c9a3/FoundationOne_Heme_Overview.pdf (accessed 07-02-22).
23. FoundationOne® Liquid CDx | Foundation Medicine. *FoundationOne Liquid CDx Technical Specifications (Gene List)*. FoundationMedicine®, 2021. https://assets.ctfassets.net/w98cd481qyp0/wVEm7VtICYR0sT5C1VbU7/fd055e0476183a6acd4eae6b583e3a00/F1LCDx_Technical_Specs_072021.pdf (accessed 07-02-22).
24. FoundationOne® CDx | Foundation Medicine. *FoundationOne CDx Technical Specifications (Gene List)*. FoundationMedicine®, 2022. https://assets.ctfassets.net/w98cd481qyp0/4mnCQaQuXydMHziIjsnukz/45a431ee91fd81e8090c00f9e8a1c9a3/FoundationOne_CDx_Technical_Specs_072021.pdf

- yp0/YqqKHaqQmFeqc5ueQk48w/d12f19680205941ea3fe417f08e9524/F1CDx_Technical_Specifications.pdf (accessed 07-02-22).
25. FoundationOne® Heme | Foundation Medicine. *FoundationOne Heme Technical Specifications (Gene List)*. Foundation Medicine®, 2021. https://assets.ctfassets.net/w98cd481qyp0/42r1cTE8VR4137CaHrsae/a7e4b78c5ee67503ede12ef0c3507de3/FoundationOne_Heme_Technical_Specifications.pdf
 26. *Target Enrichment | NGS-based hybrid capture method*. <https://www.illumina.com/techniques/sequencing/dna-sequencing/targeted-resequencing/target-enrichment.html> (accessed 2022-06-17).
 27. *Microsatellite*. <https://www.genome.gov/genetics-glossary/Microsatellite#:~:text=Microsatellite%2C%20as%20related%20to%20genomics,sequences%20are%20typically%20non%2Dcoding> (accessed 2022-06-30).
 28. Eso, Y.; Shimizu, T.; Takeda, H.; Takai, A.; Marusawa, H. Microsatellite Instability and Immune Checkpoint Inhibitors: Toward Precision Medicine against Gastrointestinal and Hepatobiliary Cancers. *Journal of Gastroenterology* 2019, 55 (1), 15–26.
 29. *Lynch syndrome: Medline Plus Genetics*. <https://medline-plus.gov/genetics/condition/lynch-syndrome/#:~:text=Lynch%20syndrome%2C%20often%20called%20hereditary,referring%20to%20as%20colorectal%20cancer> (accessed 2022-06-30).
 30. *Microsatellite Instability (MSI) screening* | CDC. https://www.cdc.gov/genomics/disease/colorectal_cancer/MSI.htm (accessed 2022-06-30).
 31. Shao, C.; Li, G.; Huang, L.; Pruitt, S.; Castellanos, E.; Frampton, G.; Carson, K. R.; Snow, T.; Singal, G.; Fabrizio, D.; Alexander, B. M.; Jin, F.; Zhou, W. Prevalence of High Tumor Mutational Burden and Association with Survival in Patients with Less Common Solid Tumors. *JAMA Network Open* 2020, 3 (10).
 32. Pardoll, D. M. The blockade of immune checkpoints in cancer immunotherapy. *Nature Reviews Cancer* 2012, 12 (4), 252–264.
 33. Duffy, M. J.; Crown, J. Biomarkers for Predicting Response to Immunotherapy with Immune Checkpoint Inhibitors in Cancer Patients. *Clinical Chemistry* 2019, 65 (10), 1228–1238.
 34. *Immune Checkpoint Inhibitors - NCI*. <https://www.cancer.gov/about-cancer/treatment/types/immunotherapy/checkpoint-inhibitors> (accessed 2022-06-30).
 35. Morjaria, S. Driver Mutations in Oncogenesis. *International Journal of Molecular and Immuno Oncology* 2020, 6 (2), 100–102.
 36. Taylor, M. R.; Simon, E. J.; Dickey, J. L.; Hogan, K. A.; Reece, J. B. How Genes Are Controlled. In *Campbell Biology: Concepts & Connections*, 9th ed.; Pearson: New York, 2018; pp 212–233.
 37. Graham, V. W. J.; Putnam, C. D.; Kolodner, R. D. DNA Mismatch Repair: Mechanisms and Cancer Genetics. In *Encyclopedia of Cancer* (Third Edition), 3rd ed.; Academic Press, 2019; pp 530–538.
 38. *DNA Is Constantly Changing through the Process of Mutation*. <https://www.nature.com/scitable/topicpage/dna-is-constantly-changing-through-the-process-6524898/#:~:text=Base%20substitution,place%20of%20a%20guanine%20nucleotide> (accessed 2022-07-03).
 39. Bacolla, A.; Cooper, D. N.; Vasquez K. M. Mechanisms of Base Substitution Mutagenesis in Cancer Genomes. *Genes* 2014, 5, 108–146.
 40. Kulkarni, S.; Pfeifer, J. D.; Kulkarni, S. Chapter 9 - Insertions and Deletions (Indels). In *Clinical Genomics*; Academic Press, 2015; pp 129–150.
 41. *Frameshift Mutation*. <https://www.genome.gov/genetics-glossary/Frameshift-Mutation#:~:text=A%20frameshift%20mutation%20in%20a,bases%20when%20making%20a%20protein> (accessed 2022-07-03).
 42. Yang, H.; Zhong, Y.; Peng, C. Important role of indels in somatic mutations of human cancer genes. *BMC Medical Genetics* 2010, 11.
 43. Esteves, L.; Caramelo, F.; Ribeiro, I. P.; Carreira, I. M.; de Melo, J. B. Probability Distribution of Copy Number Alterations along the Genome: An Algorithm to Distinguish Different Tumour Profiles. *Scientific Reports* 2020, 10.
 44. Harbers, L.; Agostini, F.; Nicos, M.; Poddighe, D.; Bienko, M.; Crosetto, N. Somatic Copy Number Alterations in Human Cancers: An Analysis of Publicly Available Data from the Cancer Genome Atlas. *Frontiers in Oncology* 2021, 11.
 45. Gu, W.; Zhang, F.; Lupski, J. R. Mechanisms for Human Genomic Rearrangements. *PathoGenetics* 2008, 1 (1).
 46. Latysheva, N. S.; Babu, M. M. Discovering and Understanding Oncogenic Gene Fusions through Data Intensive Computational Approaches. *Nucleic Acids Research* 2016, 44 (10), 4487–4503.
 47. *ROS1 and Lung Cancer* | American Lung Association. <https://lung.org/lung-health-diseases/lung-disease-lookup/lung-cancer/symptoms-diagnosis/biomarker-testing/ros1#:~:text=In%20ROS1-positive%20lung%20cancer,fuse%20with%20many%20different%20partners> (accessed 07-29-22).
 48. Chevallier, M.; Borgeaud, M.; Addeo, A.; Friedlaender, A. Oncogenic Driver Mutations in Non-Small Cell Lung Cancer: Past, Present and Future. *World Journal of Clinical Oncology* 2021, 12 (4), 217–237.
 49. *EGFR gene: MedlinePlus Genetics*. <https://medlineplus.gov/genetics/gene/egfr/> (accessed 08-03-22).
 50. *KRAS gene: MedlinePlus Genetics*. <https://medlineplus.gov/genetics/gene/kras/> (accessed 08-03-22).
 51. Referenced with permission from the NCCN Clinical Practice Guidelines in Oncology (NCCN Guidelines®) for Non-Small Cell Lung Cancer V.3.2022. © National Comprehensive Cancer Network, Inc. All rights reserved. Accessed July 28, 2022. To view the most recent and complete version of the guideline, go online to NCCN.org.
 52. Skoulidis, F.; Li, B. T.; Dy, G. K.; Price, T. J.; Falchook, G. S.; Wolf, J.; Italiano, A.; Schuler, M.; Borghaei, H.; Barlesi, F.; Kato, T.; Curioni-Fontecedro, A.; Sacher, A.; Spira, A.; Ramalingam, S. S.; Takahashi, T.; Besse, B.; Anderson, A.; Ang, A.; Tran, Q.; Mather, O.; Henary, H.; Ngarmchamnanrith, G.; Friberg, G.; Velcheti, V.; Govindan, R. Sotorasib

for Lung Cancers with KRAS P.G12C Mutation. *New England Journal of Medicine* 2021, 384 (25), 2371–2381.

53. *KRAS and Lung Cancer* | American Lung Association. [https://www.lung.org/lung-health-diseases/lung-disease-lookup/lung-cancer/symptoms-diagnosis/bio marker-testing/kRAS](https://www.lung.org/lung-health-diseases/lung-disease-lookup/lung-cancer/symptoms-diagnosis/bio-marker-testing/kRAS) (accessed 07-29-22).

54. Food and Drug Administration. *Highlights of Prescribing Information. Indications and Usage*, last revised June 2020. https://www.accessdata.fda.gov/drugsatfda_docs/label/2020/125514s084lbl.pdf (accessed 07-23-22).

55. Leonetti, A.; Sharma, S.; Minari, R.; Perego, P.; Giovannetti, E.; Tiseo, M. Resistance Mechanisms to Osimertinib EGFR-Mutated Non-Small Cell Lung Cancer. *British Journal of Cancer* 2019, 121 (9), 725–737.

56. Tian, Z.; Niu, X.; Yao, W. Receptor Tyrosine Kinases in Osteosarcoma Treatment: Which Is the Key Target? *Frontiers in Oncology* 2020, 10.

57. *Osimertinib Improves Survival in Advanced Lung Cancer with EGFR Mutations* - NCI. <https://www.cancer.gov/news-events/cancer-currents-blog/2019/osimertinib-lung-cancer-improves-survival-flaura> (accessed 2022-07-31).

58. Majeed, U.; Manochakian, R.; Zhao, Y.; Lou, Y. Targeted Therapy in Advanced Non-Small Cell Lung Cancer: Current Advances and Future Trends. *Journal of Hematology & Oncology* 2021, 14 (1).

59. *Osimertinib: MedlinePlus Drug Information*. <https://medlineplus.gov/druginfo/meds/a616005.html> (accessed 07-31-22).

60. Chen, Z.; He, X. Application of Third-Generation Sequencing in Cancer Research. *Medical Review* 2021, 1 (2), 150–171.

61. *How HiFi sequencing works* - PacBio. <https://www.pacb.com/technology/hifi-sequencing/how-it-works/> (accessed 07-30-22).

62. *Product specifications*. <https://nanoporetech.com/products/specifications> (accessed 07-30-22).

63. *How nanopore sequencing works*. <https://nanoporetech.com/how-it-works> (accessed 08-07-22).

64. Sedlazeck, F. J.; Rescheneder, P.; Smolka, M.; Fang, H.; Nattestad, M.; von Haeseler, A.; Schatz, M. C. Accurate Detection of Complex Structural Variations Using Single-Molecule Sequencing. *Nature Methods* 2018, 15 (6), 461–468.

65. Li, Y.; Roberts, N. D.; Wala, J. A.; Shapira, O.; Schumacher, S. E.; Kumar, K.; Khurana, E.; Waszak, S.; Korbel, J. O.; Haber, J. E.; Imielinski, M.; Weischenfeldt, J.; Beroukhim, R.; Campbell, P. J. Patterns of Somatic Structural Variation in Human Cancer Genomes. *Nature* 2020, 578 (7793), 112–121.

66. Macintyre, G.; Ylstra, B.; Brenton, J. D. Sequencing Structural Variants in Cancer for Precision Therapeutics. *Trends in Genetics* 2016, 32 (9), 530–542.

67. Sakamoto, Y.; Xu, L.; Seki, M.; Yokoyama, T. T.; Kasahara, M.; Kashima, Y.; Ohashi, A.; Shimada, Y.; Motoi, N.; Tsuchihara, K.; Kobayashi, S. S.; Kohno, T.; Shiraishi, Y.; Suzuki, A.; Suzuki, Y. Long-Read Sequencing for Non-Small-Cell Lung Cancer Genomes. *Genome Research* 2020, 30 (9), 1243–1257.

68. Sakamoto, Y.; Sereewattanawoot, S.; Suzuki, A. A New Era of Long-Read Sequencing for Cancer Genomics. *Journal of Human Genetics* 2019, 65 (1), 3–10.

■ Authors

Aashna Soni is a high school student at BASIS Independent Fremont, Upper School. She is interested in biology, particularly how tools like machine learning can accelerate progress in areas such as cancer research. She hopes to pursue a career in biological research. In her free time, Aashna likes to code and write short stories.