International Journal of Science and Research (IJSR)
ISSN: 2319-7064
SJIF (2022): 7.942

Assessment and Application of Quality Control Strategies in Next-Generation Sequencing Data for Pediatric Cancer

Monica Rawat[1], Ajoy Kumar Thongbam[2], Rajshree Das[3]

[1]Amity Institute of Biotechnology, Amity University, Gautam Buddha Road, Noida, Uttar Pradesh-201303, India
Email: *monica.rawat.biotech[at]gmail.com*

[2]Amity Institute of Biotechnology, Amity University, Gautam Buddha Road, Noida, Uttar Pradesh-201303, India
Email: *ajthongbam[at]gmail.com*

[3]Amity Institute of Biotechnology, Amity University, Gautam Buddha Road, Noida, Uttar Pradesh-201303, India
Email: *rdas[at]amity.edu*

Abstract: ***Next-generation sequencing (NGS) has transformed pediatric oncology by enabling precise genetic profiling of various cancers, crucial for developing targeted diagnostics and treatments. However, the extensive data produced necessitates stringent quality control (QC) to ensure accuracy and reliability, pivotal for clinical implementation. This paper reviews current QC tools and bioinformatics pipelines employed in pediatric cancer NGS analysis, emphasizing the roles of tools such as FastQC, Picard, SAMtools, and GATK. These tools assess data quality at multiple stages, helping to address challenges unique to pediatric cancer NGS, such as high specificity requirements and small sample sizes. We highlight the need for further research to develop advanced bioinformatics solutions tailored for pediatric oncology and for standardizing QC protocols to enhance data reproducibility across studies. Enhanced bioinformatics tools are essential for exploiting the full potential of NGS in improving diagnostic precision and treatment customization in pediatric oncology.***

Keywords: Next-generation sequencing, pediatric oncology, quality control, bioinformatics tools, genetic profiling

1. Introduction

Pediatric cancers encompass a range of malignancies affecting children, with leukemia being the most common type overall, followed by brain tumors, neuroblastoma, Wilms tumor, and retinoblastoma in infancy [1], [2]. Pediatric cancers include hematologic, central nervous system, and extracranial solid tumors, which are further categorized into mesenchymal-derived sarcomas and embryonal or epithelial-type tumors that often require molecular testing for diagnosis [3]. Leukemia is the most common pediatric cancer, while central nervous system tumors are the most frequent solid tumors [4]. Leukemia is the most common, with improved survival rates due to early diagnosis and new treatments [1].

Pediatric cancer is now considered a chronic condition with an 80% survival rate in the USA (United States of America) [2], [5]. Psychosocial needs evolve throughout the illness trajectory, with psychologists playing crucial roles in support and advocacy [5]. Pediatric cancers in the United States primarily consist of leukemias, lymphomas, and tumors of the central and sympathetic nervous systems, affecting over 8,500 children annually[2]. Childhood cancer incidence globally was around 200,000 in 2018 [4]. Childhood cancers have an annual incidence of 150 per million children under 15, with leukemia and CNS tumors being the most common types. Survival rates exceed 75% in industrialized countries [6].

Challenges include late diagnosis, unique immune responses, and long-term effects on growing bodies and fertility [7]. A method of treating pediatric cancers involves administering (S)-N-(5-((R)-2-(2,5-difluorophenyl)pyrrolidin-1-yl)-pyrazolo[1,5-a]pyrimidin-3-yl)-3-hydroxypyrrolidine-1-carboxamide or its salt to the patient, as per the research [8].

Worldwide efforts are being made to unravel the complexities of pediatric cancer. Efforts towards improving outcomes for pediatric cancers have been significant, with initiatives like the Childhood Cancer Data Initiative (CCDI) focusing on data-driven research to enhance understanding, survivorship, and therapy development for children and young adults with cancer [9]. Additionally, the Pediatric Cancer Dependencies Accelerator project aims to bridge knowledge gaps and enhance treatment effectiveness through a collaborative $60 million effort by leading research institutions [10]. The World Health Organization's Global Initiative for Childhood Cancer (GICC) is also playing a crucial role in improving pediatric cancer services in Africa by implementing the CureAll framework, emphasizing essential pillars like adequately staffed care networks, universal health coverage, context-appropriate treatments, and evaluation strategies [11]. Furthermore, regulatory decisions on pediatric investigation plans (PIPs) in oncology are evolving to support the development of innovative and safe medicines for children with high unmet medical needs, highlighting the importance of evidence-based decision-making and stakeholder engagement [12].

Various types of research have been conducted to identify genes related to pediatric cancer. Pedican is a literature-based gene resource for pediatric cancers, containing 735 genes, 88 gene fusions, and 24 chromosome abnormalities curated from 2245 PubMed abstracts [13]. Studies have focused on the genetic basis of childhood cancers, highlighting the prevalence of germline alterations in cancer predisposition genes and the heterogeneity of genetic alterations driving

tumorigenesis in pediatric cancers [14]. Efforts have been made to develop tools for identifying children at increased risk of cancer through genetic testing, emphasizing the need for validation and large-scale application of these tools. Research on the genetic bases of diseases using various sequencing approaches is crucial for identifying high-risk patients who may benefit from genetic testing [15].

In pediatric cancer research, various sequencing approaches are utilized to enhance diagnosis and treatment. Studies have employed low-pass whole genome sequencing (LPWGS) and targeted sequencing (TS) of cell-free DNA to detect copy number alterations, fusions, and driver mutations in pediatric solid and brain tumors, showcasing the potential of liquid biopsy assays for diagnosis and minimal residual disease monitoring [16]. Additionally, RNA-Seq assays have been developed to identify gene fusions, expressed mutations, and gene expression abnormalities in pediatric tumors, demonstrating the utility of tumor RNA profiling in clinical settings [17]. Furthermore, targeted genomic sequencing approaches have been explored to overcome the challenges of lower mutation frequency in pediatric cancers, emphasizing the importance of clinically available and targeted sequencing methods for tumor genomic profiling in pediatric oncology [18]. Moreover, the implementation of next-generation sequencing (NGS) panels in middle-income countries like Mexico has shown promise in improving diagnostic accuracy and identifying actionable genetic alterations for risk stratification and tailored therapy in pediatric cancer patients [19]. These sequencing strategies collectively contribute to advancing precision medicine and personalized treatment approaches in pediatric oncology, aiming to improve outcomes for young cancer patients [20].

Next-generation sequencing (NGS) approaches used in pediatric cancer include a three-platform sequencing strategy incorporating whole-genome sequencing (WGS), whole-exome sequencing (WES), and RNA sequencing (RNA-seq) to analyze tumor and germline genomes revealing diagnostic, prognostic, therapeutically relevant, and cancer-predisposing variants [21], [22]. Additionally, low-pass whole genome sequencing (LPWGS) and targeted sequencing (TS) of cell-free DNA (cfDNA) in pediatric solid and brain tumors have shown promise in detecting aberrations, with a high positivity rate in treatment-naive cases and potential for minimal residual disease (MRD) monitoring [23].

Whole-genome sequencing (WGS) is paramount in pediatric oncology as it yields diagnostically significant data across diverse patient populations, thus facilitating actionable insights for therapeutic strategies [24]. Complementing WGS with whole-transcriptome sequencing (RNA-Seq) notably enhances fusion gene detection, which is crucial for the accurate subclassification of tumors and pinpointing prognostic indicators [24] . The integration of WGS and RNA-Seq enables the identification of high confidence tumor-specific gene fusions, potentially pathogenic fusions, and associated structural variants, aiding in precision oncology applications and clinical decision-making[25].

Whole-exome sequencing (WES) for pediatric cancer is highly relevant as it enables the comprehensive examination of genetic alterations in both tumor and germline genomes, providing crucial diagnostic, prognostic, and therapeutic insights. Studies have shown that WES, when combined with whole-genome sequencing (WGS) and RNA sequencing (RNA-seq), allows for the identification of diverse genomic variants in pediatric cancers, regardless of tumor type, with a high percentage of patients harboring clinically actionable variants[24]. WES has been instrumental in detecting activating gene fusions, enhancer hijacks, small intragenic deletions, and mutational signatures, shedding light on the pathogenic effects of variants and their relevance to tumor development[22]. Additionally, the integration of WES data in multidisciplinary molecular tumor boards facilitates informed clinical decision-making, contributing to improved patient management and outcomes in pediatric oncology [22].

RNA sequencing (RNA-seq) is highly relevant for pediatric cancer as it enables the detection of gene fusions, expressed mutations, and important gene expression abnormalities in tumors, aiding in diagnosis and treatment decisions[25]. By integrating RNA-seq with whole genome sequencing (WGS), Fusion-sq can identify tumor-specific gene fusions with high confidence, distinguishing them from healthy-occurring fusions and resolving fusions in copy number unstable genomes[17]. A three-platform sequencing approach incorporating WGS, whole-exome sequencing (WES), and RNA-seq has shown the power to identify diverse genomic lesions in pediatric cancers, enabling the detection of diagnostic, prognostic, and therapeutically relevant variants, highlighting the importance of RNA-seq in understanding the genomic landscape of pediatric cancers [26]. Fusion-sq is a novel method developed to detect tumor-specific gene fusions in pediatric cancers by integrating evidence from RNA sequencing (RNA-seq) and whole genome sequencing (WGS) using intron-exon gene structure [25]. Next-generation sequencing (NGS) technologies, including whole-genome sequencing (WGS), whole-exome sequencing (WES), and RNA-seq, have revolutionized genomics but come with challenges. The vast amount of data generated by NGS requires sophisticated bioinformatics tools for analysis and interpretation [27], [28]. Identifying causative variants in rare diseases is complex due to the increased number of variants to evaluate per patient, leading to diagnostic odysseys and a rise in variants of uncertain significance (VUSs)[29]. In the clinical setting, challenges include the need for precise sequencing, data curation, and accurate genome annotation to derive meaningful insights for disease prognosis, diagnosis, and treatment [30], [31]

2. Challenges in NGS for Pediatric Cancer

Achieving high specificity and sensitivity in next-generation sequencing (NGS) analyses for pediatric tumors presents significant challenges due to diverse molecular features, limited sample sizes, and the need for precise variant interpretation. Variability in technically challenging variants, such as large indels and small copy number variations, can impact the diagnostic yield and clinical sensitivity of NGS workflows, affecting test design and validation for pediatric cancers. Additionally, the rarity of pediatric cancers and heterogeneity within narrow types pose difficulties in risk stratification and treatment selection, emphasizing the need for accurate variant classification and therapy matching. Sequencing methods also face obstacles such as the limited

availability of targeted drugs due to low mutation frequency and the use of large-scale genome-wide screening applications, which hinder the clinical integration of tumor genomic profiling in pediatric cancers [18]. Other challenges include limited amounts of cell-free DNA (cfDNA), a higher prevalence of copy number alterations and fusions over point mutations, and small sample sizes, impacting the progress of liquid biopsy assays for diagnosis and monitoring [16]. Moreover, extracting sufficient mRNA from children with low white blood cell counts complicates the diagnosis of infectious diseases [26], while the reliance on multiple tests like FISH, karyotyping, microarray, and sequencing for genetic characterization leads to longer analysis times, increased costs, and higher consumption of limited tumor material [32]. Efforts to implement genomic sequencing in low- and middle-income countries are further challenged by human resource training needs and site-level implementation costs [33], to address these a set of specialized algorithms and enhanced sample preparation techniques are employed. The FDA-led SEQC2 project focuses on developing standard analysis protocols and quality control metrics for DNA testing, enhancing precision medicine [34], [35].

Specialized algorithms significantly enhance next-generation sequencing (NGS) quality control by providing accurate and automated procedures. They facilitate the detection of erroneous base calls, removal of contaminating sequences, and mitigation of issues like barcode swapping, ensuring high-quality data output [36]. These specialized algorithms serve as valuable tools for NGS specialists, enabling them to better understand quality issues, perform automatic quality control, and improve the reliability and efficiency of NGS experiments. Additionally, the AmpliSeq for Illumina Childhood Cancer Panel demonstrates high sensitivity, specificity, and reproducibility, refining pediatric acute leukemia diagnosis, prognosis, and treatment through a targeted NGS panel [37]. These approaches collectively contribute to improving the reliability and clinical utility of NGS in pediatric oncology [37].

3. Quality Control Metrics and Standards

In pediatric cancer sequencing, QC metrics play a crucial role in ensuring the accuracy and reliability of the results. Various metrics such as coverage depth, error rate, and alignment quality are essential for assessing the quality of sequencing data [38]. Studies have shown that analytical validation of RNA-Seq assays in pediatric oncology clinics involves evaluating fusion detection with high precision, establishing the limit of detection for gene fusions, and assessing gene expression measurements for accuracy[17], [39]. Additionally, a three-platform sequencing approach incorporating WGS, WES, and RNA-seq has been utilized to identify diagnostic, prognostic, and therapeutically relevant variants in pediatric cancers, demonstrating the importance of comprehensive genomic analysis for understanding the full spectrum of genomic variants in pediatric oncology[22]. These QC metrics are vital for ensuring the robustness and validity of sequencing data, ultimately leading to more accurate diagnoses and targeted treatment strategies for pediatric cancer patients [38].

These advancements in NGS technology not only refine diagnostic and prognostic capabilities but also pave the way for more precise and effective therapeutic interventions in pediatric oncology, ultimately improving patient outcomes and the overall management of childhood cancer [40]. Efforts to standardize these metrics across laboratories are ongoing to promote consistency and reliability in NGS data analysis. Studies emphasize the importance of analytical validation to establish high accuracy, sensitivity, repeatability, and reproducibility in detecting genomic alterations, thus enabling comprehensive molecular profiling of pediatric cancers with fast turnaround times and high-quality results[24], [41]. Standardization initiatives aim to ensure that QC metrics are uniformly applied, enhancing the interpretation and clinical utility of genomic data in pediatric oncology [42].

4. Review of Quality Control Tools

4.1 Pre-Sequencing QC Tools

In the realm of next-generation sequencing (NGS), quality control (QC) initiates even before sequencing commences. Sample Integrity Checks are vital to verify the integrity and purity of DNA or RNA extracted from samples, ensuring reliable sequencing outcomes [43]. Subsequently, Library Preparation QC plays a crucial role in evaluating the quality and size of the library, guaranteeing adequate representation of the sample [42]. These pre-sequencing QC steps are fundamental in laying the groundwork for a successful sequencing run, setting the stage for accurate and meaningful downstream analyses. Various tools and technologies have been developed to streamline and standardize these QC processes, enhancing the efficiency and reliability of NGS workflows [44], [45], [46].

4.2 Sequencing Run QC

Real-time monitoring during a sequencing run is crucial for tracking the process and promptly addressing any emerging issues like signal intensity drops or base calling errors, as highlighted in the research papers [42], [47]. This real-time analysis is essential for optimizing flow cell performance, balancing coverage across samples, and reducing sequencing time and costs, especially in amplicon-based sequencing scenarios [48].

4.3 Post- sequencing data analysis

Post-sequencing data analysis methods for pediatric cancer analysis involve various approaches. These include exploring germline secondary findings (SFs) through whole-genome sequencing to identify pathogenic variants and classify their significance[49]. Additionally, RNA editing events in pediatric cancers are investigated using RNA-Seq data to understand protein sequence diversity and biological implications [50]. A cancer genomic analysis pipeline has been developed to analyze genetic mutations in pediatric tumors, highlighting driver mutations and their impact on gene expression [51]. Liquid biopsy analysis for pediatric tumors utilizes deep whole-genome sequencing profiles to detect circulating tumor DNA and classify cancer-specific chromatin signatures, providing insights into minimally

invasive disease monitoring and treatment response [52]. Furthermore, the development of an RNA-Seq assay pipeline aims to identify gene fusions, expressed mutations, and gene expression abnormalities in pediatric tumors, advancing molecular diagnostic methods in pediatric oncology [17].

5. Tools used for Post-Sequencing Quality Check

In the realm of pediatric oncology, the integrity of sequences in NGS post-processing is evaluated through a variety of quality metrics. These metrics include measuring the mean read depth, sensitivity for DNA and RNA variants, specificity, reproducibility for DNA and RNA, clinical impact of mutations and fusions, refinement of diagnosis, targetability of mutations, and clinical relevance of results [37]. Additionally, metrics such as adequacy rate, initial DNA concentration, fragment size, insert size, total reads, target coverage, duplication rate, and usable bases are crucial for evaluating the quality of samples and libraries in NGS analysis [53], [54]. Furthermore, for multi-sample experiments like tumor-normal pairs, specific QC metrics are essential to ensure the validity of results, with proposed workflows and tools available for comprehensive QC analysis [55]. An overview of the tools is given in Table 1.

Table 1: Overview of Quality Control Tools

Feature/Tool	Fast QC	SAM tools	GATK
Primary Function	Initial quality assessment of raw sequencing data.	Managing SAM/BAM files including duplicate marking.	Comprehensive genomic data analysis including variant calling and quality recalibration.
Quality Metrics	- Per Base Sequence Quality - Per Sequence Quality Scores	- Read and Mate Reference Positions - Pair Orientation	- Alignment Metrics - Variant Calling Metrics
Content Analysis	- Per Base Sequence Content - Per Sequence GC Content	N/A	N/A
Base/Sequence Specific Checks	- Per Base N Content - Sequence Length Distribution	- Leftmost Position	- Haplotype Caller
Duplication and Overrepresentation	- Sequence Duplication Levels - Overrepresented Sequences	- UMI/Barcodes (optional) - Read Groups (optional)	N/A
Adapter and Contaminant Screening	- Adapter Content	N/A	N/A
Quality Adjustments and Filtering	N/A	- Base Quality Scores	- Base Quality Score Recalibration (BQSR) - Variant Filtration

5.1 FASTQC

FastQC is a widely used tool for quality control in DNA sequencing analysis, deeply integrated into standard pipelines at sequencing centers [56]. It plays a crucial role in the initial stages of data analysis by providing detailed diagnostic information on sequencing data quality, highlighting potential issues such as per base sequence content, per tile sequence quality, and adapter content warnings[57]. FastQC is particularly valuable for assessing the quality of DNA samples, confirming their origin through mapping reads to reference genomes[58]. Additionally, FastQC aids in optimizing computing resources by offering faster computation times and reduced memory requirements compared to other tools like Falco, ensuring efficient quality control processes in sequencing data analysis[59]. Overall, FastQC is essential for ensuring the reliability and accuracy of sequencing data, especially in scenarios where multiple genomes need to be considered or when uncertainty exists regarding the DNA sample's origin[60].

FastQC provides several metrics to ensure the quality of high-throughput sequencing data:

1) Per Base Sequence Quality: Visualizes the quality score across each base position in the reads.
2) Per Sequence Quality Scores: Shows the distribution of quality scores across all reads.
3) Per Base Sequence Content: Assesses base composition across each position.
4) Per Sequence GC Content: Analyzes the GC distribution in the reads compared to a theoretical distribution.
5) Per Base N Content: Indicates the count of bases called as 'N' per position.
6) Sequence Length Distribution: Checks if all the sequences are of uniform length.
7) Sequence Duplication Levels: Highlights the degree of duplication.
8) Overrepresented Sequences: Identifies sequences that appear more often than statistically expected.
9) Adapter Content: Detects the presence of adapter sequences in the data.[61]

5.2 SAM tools

SAMtools is a crucial tool for quality control in next-generation sequencing (NGS) bioinformatics analysis, especially in pediatric cancer research. It aids in excluding duplicates and ensuring the accuracy of variant calling [62]. Quality control (QC) is essential in all stages of NGS data analysis, with tools like SAMtools providing metrics for single-sample experiments and tumor-normal pairs [63]. Additionally, a quality control system has been developed to enhance the NGS detection process in tumor samples, ensuring comprehensive quality control from experiment source to quantification results, thereby improving experiment efficiency and reliability [55]. By utilizing SAMtools and other QC tools, researchers can maintain data integrity, enhance the accuracy of variant calling, and ensure the reliability of NGS data analysis in pediatric cancer studies. Samtools uses the following metrics to ensure quality through duplicate marking:

1) Read and Mate Reference Positions: Evaluates if reads align to the same reference and position.

2) Pair Orientation: Checks the orientation of the read pairs.
3) Leftmost Position: Determines if the read is the leftmost (lowest position) in a pair.
4) UMI/Barcodes (optional): Adds barcode checking to further identify duplicates.
5) Read Groups (optional): Uses read group information from the RG tag to distinguish duplicates.
6) Base Quality Scores: Uses the sum of base quality scores to identify the original among duplicates.[64]

5.3 GATK (Genome Analysis ToolKit)

The Genome Analysis Toolkit (GATK) provides a robust framework for quality control (QC) tools in next-generation sequencing (NGS) bioinformatics analysis, particularly in pediatric cancer research. GATK's structured programming philosophy facilitates the development of efficient and reliable NGS tools, including coverage calculators and single nucleotide polymorphism (SNP) calling, essential for analyzing genomic variants in cancer studies [65]. Additionally, the K-mer Analysis Toolkit (KAT) offers reference-free QC for WGS reads and de novo genome assemblies, aiding in error detection and assembly quality assessment [66]. Furthermore, the NGS QC Toolkit provides user-friendly tools for QC and filtering of high-quality data generated from NGS platforms like Illumina, crucial for downstream analysis in pediatric cancer research [67]. Integrating these tools can enhance the precision and efficiency of NGS bioinformatics analysis in pediatric cancer studies, ensuring accurate and meaningful conclusions from the rich cancer genomics data available [68].

The GATK (Genome Analysis Toolkit) provides several metrics and tools for ensuring high-quality genomic data analysis as mentioned below:

1) Alignment Metrics: Assess the quality and efficiency of alignment to the reference genome.
2) Base Quality Score Recalibration (BQSR): Adjusts the base quality scores of sequencing reads to correct for biases that arise during the sequencing process.
3) Variant Calling Metrics: Evaluate the quality of variant calls, considering factors like depth of coverage and quality by depth.
4) HaplotypeCaller: Generates accurate variant calls with evaluation of evidence supporting each possible haplotype.
5) Variant Filtration: Applies filters to variant calls based on quality metrics to ensure the reliability of the results.
6) Depth of Coverage: Analyzes the depth of coverage across the genome to identify areas with low coverage that may affect variant calling accuracy.

These metrics and tools are part of a comprehensive approach to optimize the accuracy and reliability of NGS data analysis using GATK.

6. Conclusion

The ongoing evolution of next-generation sequencing (NGS) technologies has catalyzed significant advancements in pediatric cancer research, offering insights into the complex genomic landscapes of various malignancies. However, the vast data generated by NGS requires rigorous quality control (QC) measures to ensure the accuracy and reliability of the findings, which are necessary for clinical decision-making. The current QC tools, while effective, underscore the necessity for continual development of more sophisticated bioinformatics pipelines. These enhancements are essential to manage the intricacies of pediatric cancer genomics, where the accuracy of variant detection and interpretation directly impacts therapeutic outcomes.

Future research in NGS bioinformatics should focus on refining algorithms that enhance the sensitivity and specificity of data analysis, integrating artificial intelligence to handle data complexity, and developing standardized protocols that can be universally applied to improve reproducibility across studies. Moreover, as pediatric cancers are notably diverse and often present unique genomic signatures compared to adult cancers, there is a pressing need for tailored bioinformatics approaches that address these specific challenges.

Emphasizing the development of robust, scalable, and flexible bioinformatics solutions will not only streamline diagnostic workflows but also expedite the translation of genomic discoveries into clinical applications. This will ultimately foster personalized treatment strategies that are more effective and less invasive, marking a significant step forward in the fight against pediatric cancer. Thus, continued investment in research and development within this circle is imperative to harness the full potential of NGS technologies, aiming to transform pediatric oncology by enhancing diagnostic precision and therapeutic efficacy.

References

[1] S. Mohiuddin, W. Zaky, and J. Cortes, ‘Overview of Pediatric Cancers’, in *Perioperative Care of the Cancer Patient*, Elsevier, 2023, pp. 491–497. doi: 10.1016/B978-0-323-69584-8.00045-1.
[2] J. A. Ross and A. F. Olshan, ‘Pediatric Cancer in the United States: The Children’s Oncology Group Epidemiology Research Program’, *Cancer Epidemiology, Biomarkers & Prevention*, vol. 13, no. 10, pp. 1552–1554, Oct. 2004, doi: 10.1158/1055-9965.1552.13.10.
[3] M. Warren, ‘Pediatric solid tumors’, in *Cases in Laboratory Genetics and Genomics (LGG) Practice*, Elsevier, 2023, pp. 427–477. doi: 10.1016/B978-0-323-99622-8.00029-0.
[4] C. Stiller, ‘Epidemiology of Pediatric Cancer’, in *Pediatric Surgical Oncology*, 1st ed., Boca Raton: CRC Press, 2022, pp. 3–9. doi: 10.1201/9781351166126-2.
[5] K. R. Parris, J. M. Allen, and N. Jurbergs, ‘Pediatric Cancer’, in *Handbook of Clinical Child Psychology*, J. L. Matson, Ed., in Autism and Child Psychopathology Series. , Cham: Springer International Publishing, 2023, pp. 745–762. doi: 10.1007/978-3-031-24926-6_34.
[6] C. A. Stiller, ‘Epidemiology of Childhood Tumours’, in *The Surgery of Childhood Tumors*, R. Carachi and J. L. Grosfeld, Eds., Berlin, Heidelberg: Springer Berlin Heidelberg, 2016, pp. 7–18. doi: 10.1007/978-3-662-48590-3_2.

[7] S. Agrawal and N. Bansal, 'Advances and challenges in pediatric cancers', *Cancer Reports*, vol. 2, no. 3, p. e1202, Jun. 2019, doi: 10.1002/cnr2.1202.
[8] M. Cox and N. Nanda, 'Methods of treating pediatric cancers', US10045991B2, Aug. 14, 2018 Accessed: Jun. 14, 2024. [Online]. Available: https://patents.google.com/patent/US10045991B2/en
[9] J. A. Flores-Toro *et al.*, 'The Childhood Cancer Data Initiative: Using the Power of Data to Learn From and Improve Outcomes for Every Child and Young Adult With Pediatric Cancer', *JCO*, vol. 41, no. 24, pp. 4045–4053, Aug. 2023, doi: 10.1200/JCO.22.02208.
[10] 'Institutions Launch $60 Million Effort to Improve Kids' Cancer Care', *Cancer Discovery*, vol. 13, no. 1, pp. OF5–OF5, Jan. 2023, doi: 10.1158/2159-8290.CD-NB2022-0068.
[11] J. Balagadde-Kambugu *et al.*, 'Toward 2030: SIOP Africa adopts the Global Initiative for Childhood Cancer', *Pediatric Hematology and Oncology*, vol. 40, no. 3, pp. 197–202, Apr. 2023, doi: 10.1080/08880018.2022.2117883.
[12] D. Karres *et al.*, 'European regulatory strategy for supporting childhood cancer therapy developments', *European Journal of Cancer*, vol. 177, pp. 25–29, Dec. 2022, doi: 10.1016/j.ejca.2022.09.025.
[13] M. Zhao, L. Ma, Y. Liu, and H. Qu, 'Pedican: an online gene resource for pediatric cancers with literature evidence', *Sci Rep*, vol. 5, no. 1, p. 11435, Jun. 2015, doi: 10.1038/srep11435.
[14] A. Filipiuk, A. Kozakiewicz, K. Kośmider, M. Lejman, and J. Zawitkowska, 'Genetic Disorders with Predisposition to Paediatric Haematopoietic Malignancies—A Review', *Cancers*, vol. 14, no. 15, Art. no. 15, Jan. 2022, doi: 10.3390/cancers14153569.
[15] O. Alonso-Luna, G. E. Mercado-Celis, J. Melendez-Zajgla, M. Zapata-Tarres, and E. Mendoza-Caamal, 'The genetic era of childhood cancer: Identification of high-risk patients and germline sequencing approaches', *Annals of Human Genetics*, vol. 87, no. 3, pp. 81–90, May 2023, doi: 10.1111/ahg.12502.
[16] V. Yellapantula *et al.*, 'Abstract 3322: Diagnosis and monitoring of pediatric cancer patients using low-pass whole genome and targeted sequencing of cell-free DNA', *Cancer Research*, vol. 83, no. 7_Supplement, pp. 3322–3322, Apr. 2023, doi: 10.1158/1538-7445.AM2023-3322.
[17] O. M. Vaske *et al.*, 'Comprehensive clinical tumor RNA sequencing for children with cancer.', *JCO*, vol. 41, no. 16_suppl, pp. e15067–e15067, Jun. 2023, doi: 10.1200/JCO.2023.41.16_suppl.e15067.
[18] G. M. Turco *et al.*, 'An institutional review of genomic sequencing in pediatric solid tumors', *Pediatric Blood & Cancer*, vol. 70, no. 6, p. e30324, Jun. 2023, doi: 10.1002/pbc.30324.
[19] R. Ruiz-Cordero *et al.*, 'Abstract 6: Applying Next-Generation Sequencing to Improve Pediatric Cancer Diagnosis in Mexico', *Cancer Epidemiology, Biomarkers & Prevention*, vol. 32, no. 6_Supplement, pp. 6–6, Jun. 2023, doi: 10.1158/1538-7755.ASGCR23-Abstract-6.
[20] K. V. Okonechnikov, M. V. Ryzhova, S. A. Galstyan, and E. N. Telysheva, '[Application of high throughput sequencing in pediatric neurooncology]', *Arkh Patol*, vol. 84, no. 2, pp. 58–63, 2022, doi: 10.17116/patol20228402158.
[21] J. Y. Choi *et al.*, 'Abstract 3398: Clinical use of next-generation sequencing panel in pediatric oncology patients', *Cancer Research*, vol. 83, no. 7_Supplement, pp. 3398–3398, Apr. 2023, doi: 10.1158/1538-7445.AM2023-3398.
[22] S. Newman *et al.*, 'Data from Genomes for Kids: The Scope of Pathogenic Mutations in Pediatric Cancer Revealed by Comprehensive DNA and RNA Sequencing', American Association for Cancer Research. Accessed: Jun. 14, 2024. [Online]. Available: https://aacr.figshare.com/collections/Data_from_Genomes_for_Kids_The_Scope_of_Pathogenic_Mutations_in_Pediatric_Cancer_Revealed_by_Comprehensive_DNA_and_RNA_Sequencing/6549437
[23] A. Nitzan - Luques, H. Cohen, S. Winerb, D. Raviv, D. Harlev, and G. Goldstein, 'THE BENEFIT OF MOLECULAR ANALYSIS USING NEXT GENERATION SEQUENCING METHODS IN PEDIATRIC "HARD TO TREAT" TUMORS - EXPERIENCE IN A TERTIARY MEDICAL CENTER', *Harefuah*, vol. 162, pp. 24–30, Jan. 2023.
[24] E. Wadensten *et al.*, 'Diagnostic Yield From a Nationwide Implementation of Precision Medicine for all Children With Cancer', *JCO Precision Oncology*, no. 7, p. e2300039, Jun. 2023, doi: 10.1200/PO.23.00039.
[25] I. A. E. M. Van Belzen *et al.*, 'Systematic discovery of gene fusions in pediatric cancer by integrating RNA-seq and WGS', *BMC Cancer*, vol. 23, no. 1, p. 618, Jul. 2023, doi: 10.1186/s12885-023-11054-3.
[26] V. Probst *et al.*, 'A Protocol for Low-Input RNA-Sequencing of Patients with Febrile Neutropenia Captures Relevant Immunological Information', *International Journal of Molecular Sciences*, vol. 24, no. 12, Art. no. 12, Jan. 2023, doi: 10.3390/ijms241210251.
[27] I. Mani, 'Current Status and Challenges of DNA Sequencing', in *Advances in Synthetic Biology*, V. Singh, Ed., Singapore: Springer Singapore, 2020, pp. 71–80. doi: 10.1007/978-981-15-0081-7_5.
[28] S. Tian, Z. J. Tu, H. Yan, and E. W. Klee, 'Editorial: Clinical Genome Sequencing: Bioinformatics Challenges and Key Considerations', *Front. Genet.*, vol. 13, p. 896032, Mar. 2022, doi: 10.3389/fgene.2022.896032.
[29] G. Fernandez, D. Yubero, F. Palau, and J. Armstrong, 'Molecular Modelling Hurdle in the Next-Generation Sequencing Era', *IJMS*, vol. 23, no. 13, p. 7176, Jun. 2022, doi: 10.3390/ijms23137176.
[30] S. H. Eck, 'Challenges in data storage and data management in a clinical diagnostic setting', *LaboratoriumsMedizin*, vol. 42, no. 6, pp. 219–224, Dec. 2018, doi: 10.1515/labmed-2018-0054.
[31] V. Novelli and M. Vatta, 'Editorial: Current Challenges in Cardiovascular Molecular Diagnostics', *Front. Cardiovasc. Med.*, vol. 4, p. 54, Sep. 2017, doi: 10.3389/fcvm.2017.00054.
[32] B. Yoo *et al.*, 'Abstract 243: Long-read sequencing of pediatric cancer genomes identifies multiple clinically relevant variants', *Cancer Research*, vol. 83, no.

7_Supplement, pp. 243–243, Apr. 2023, doi: 10.1158/1538-7445.AM2023-243.
[33] K. Busby *et al.*, 'Abstract 62: Minding the Gap: Implementation of Sequencing-Based Diagnostics for Pediatric Cancer in Resource-Limited Settings', *Cancer Epidemiology, Biomarkers & Prevention*, vol. 32, no. 6_Supplement, pp. 62–62, Jun. 2023, doi: 10.1158/1538-7755.ASGCR23-Abstract-62.
[34] I. Alborelli and P. M. Jermann, 'Preanalytical Variables and Sample Quality Control for Clinical Variant Analysis', in *Variant Calling*, vol. 2493, C. Ng and S. Piscuoglio, Eds., in Methods in Molecular Biology, vol. 2493. , New York, NY: Springer US, 2022, pp. 331–351. doi: 10.1007/978-1-0716-2293-3_21.
[35] J. C. Willey *et al.*, 'Advancing NGS quality control to enable measurement of actionable mutations in circulating tumor DNA', *Cell Reports Methods*, vol. 1, no. 7, p. 100106, Nov. 2021, doi: 10.1016/j.crmeth.2021.100106.
[36] R. D. Kuster, G. C. Yencho, and B. A. Olukolu, 'ngsComposer: an automated pipeline for empirically based NGS data quality filtering', *Briefings in Bioinformatics*, vol. 22, no. 5, p. bbab092, Sep. 2021, doi: 10.1093/bib/bbab092.
[37] C. Vicente-Garcés *et al.*, 'Technical Validation and Clinical Utility of an NGS Targeted Panel to Improve Molecular Characterization of Pediatric Acute Leukemia', *Front. Mol. Biosci.*, vol. 9, Apr. 2022, doi: 10.3389/fmolb.2022.854098.
[38] E. Khlebus *et al.*, 'Table S1 from Comparative Tumor Microenvironment Analysis of Primary and Recurrent Ovarian Granulosa Cell Tumors', American Association for Cancer Research. Accessed: Jun. 15, 2024. [Online]. Available: https://aacr.figshare.com/articles/dataset/Table_S1_from_Comparative_Tumor_Microenvironment_Analysis_of_Primary_and_Recurrent_Ovarian_Granulosa_Cell_Tumors/22725314/1
[39] K. F. Schilter *et al.*, 'Analytical validation and implementation of a pan cancer next-generation sequencing panel, CANSeqTMKids for molecular profiling of childhood malignancies', *Front. Genet.*, vol. 14, p. 1067457, Feb. 2023, doi: 10.3389/fgene.2023.1067457.
[40] J. C. Willey *et al.*, 'Advancing quality-control for NGS measurement of actionable mutations in circulating tumor DNA'. Apr. 06, 2021. doi: 10.1101/2021.04.06.438497.
[41] J. Anaya, J.-W. Sidhom, C. A. Cummings, and A. S. Baras, 'Supplemental Table ST2. TMB regression metrics for different ancestries. from Probabilistic Mixture Models Improve Calibration of Panel-derived Tumor Mutational Burden in the Context of both Tumor-normal and Tumor-only Sequencing', American Association for Cancer Research. Accessed: Jun. 15, 2024. [Online]. Available: https://aacr.figshare.com/articles/journal_contribution/Supplemental_Table_ST2_TMB_regression_metrics_for_different_ancestries_from_Probabilistic_Mixture_Models_Improve_Calibration_of_Panel-derived_Tumor_Mutational_Burden_in_the_Context_of_both_Tumor-normal_and_Tumor-only_Sequencing/22348724
[42] M. Gajapathy, B. M. Wilk, and E. A. Worthey, 'QuaC: A Pipeline Implementing Quality Control Best Practices for Genome Sequencing and Exome Sequencing Data'. Mar. 07, 2023. doi: 10.1101/2023.03.06.531383.
[43] A. S. C. Gaia *et al.*, 'e-QA NGS: a user-friendly tool to preprocessing data from next generation sequencing', *PRW*, vol. 5, no. 3, pp. 91–105, Mar. 2023, doi: 10.53660/201.prw203.
[44] E. Box and C. Gleasner, 'Sample Prep and Initial QC (DNA) [Slides]', LA-UR--23-25242, 1973798, May 2023. doi: 10.2172/1973798.
[45] W. Macnair and M. D. Robinson, 'SampleQC: robust multivariate, multi-celltype, multi-sample quality control for single cell data'. bioRxiv, p. 2021.08.28.458012, Apr. 05, 2022. doi: 10.1101/2021.08.28.458012.
[46] N. Roseman *et al.*, 'Abstract 263: Automating next generation sequencing workflows for custom hyb panels: high quality library preparation and target enrichment on the Biomek NGeniuS System', *Cancer Research*, vol. 83, no. 7_Supplement, pp. 263–263, Apr. 2023, doi: 10.1158/1538-7445.AM2023-263.
[47] R. Munro *et al.*, 'A framework for real-time monitoring, analysis and adaptive sampling of viral amplicon nanopore sequencing', *Front. Genet.*, vol. 14, p. 1138582, Mar. 2023, doi: 10.3389/fgene.2023.1138582.
[48] R. Bedre, C. Avila, and K. Mandadi, 'HTSQualC is a Flexible and One-Step Quality Control Software for High-throughput Sequencing Data Analysis'. Jul. 24, 2020. doi: 10.1101/2020.07.23.214536.
[49] S. Majeed *et al.*, 'Abstract 6545: Investigating secondary findings in a pediatric cancer cohort: preliminary findings', *Cancer Research*, vol. 83, no. 7_Supplement, pp. 6545–6545, Apr. 2023, doi: 10.1158/1538-7445.AM2023-6545.
[50] J. Wen *et al.*, 'The landscape of coding RNA editing events in pediatric cancer', *BMC Cancer*, vol. 21, no. 1, p. 1233, Dec. 2021, doi: 10.1186/s12885-021-08956-5.
[51] M. Zeller, C. N. Magnan, V. R. Patel, P. Rigor, L. Sender, and P. Baldi, 'A Genomic Analysis Pipeline and Its Application to Pediatric Cancers', *IEEE/ACM Trans Comput Biol Bioinform*, vol. 11, no. 5, pp. 826–839, 2014, doi: 10.1109/TCBB.2014.2330616.
[52] P. Peneder *et al.*, 'Multimodal analysis of cell-free DNA whole-genome sequencing for pediatric cancers with low mutational burden', *Nat Commun*, vol. 12, no. 1, p. 3230, May 2021, doi: 10.1038/s41467-021-23445-w.
[53] D. H. Hwang, E. P. Garcia, M. D. Ducar, E. S. Cibas, and L. M. Sholl, 'Next-generation sequencing of cytologic preparations: An analysis of quality metrics', *Cancer Cytopathology*, vol. 125, no. 10, pp. 786–794, 2017, doi: 10.1002/cncy.21897.
[54] J. Langmore, 'Quality Control and Pre-Qualification of NGS Libraries Made from Clinical Samples', *J Biomol Tech*, vol. 24, no. Suppl, p. S46, May 2013.
[55] C. M. Schroeder, F. J. Hilke, M. W. Löffler, M. Bitzer, F. Lenz, and M. Sturm, 'A comprehensive quality control workflow for paired tumor-normal NGS experiments', *Bioinformatics*, vol. 33, no. 11, pp. 1721–1722, Jun. 2017, doi: 10.1093/bioinformatics/btx032.
[56] S. Al Yami and C.-H. Huang, 'LFastqC: A lossless non-reference-based FASTQ compressor', *PLoS ONE*, vol.

14, no. 11, p. e0224806, Nov. 2019, doi: 10.1371/journal.pone.0224806.
[57] S. W. Wingett and S. Andrews, 'FastQ Screen: A tool for multi-genome mapping and quality control', *F1000Res*, vol. 7, p. 1338, Sep. 2018, doi: 10.12688/f1000research.15931.2.
[58] Honghu Quan, D. Arsala, and J. Lynch, 'FastQC results of the raw sequences from first analysis.', p. 9336528 Bytes, 2019, doi: 10.6084/M9.FIGSHARE.8146040.V1.
[59] G. De Sena Brandine and A. D. Smith, 'Falco: high-speed FastQC emulation for quality control of sequencing data', *F1000Res*, vol. 8, p. 1874, Jan. 2021, doi: 10.12688/f1000research.21142.2.
[60] S. Gómez *et al.*, 'FastIC: a fast integrated circuit for the readout of high performance detectors', *J. Inst.*, vol. 17, no. 05, p. C05027, May 2022, doi: 10.1088/1748-0221/17/05/C05027.
[61] B. B. Babraham Bioinformatics, 'Babraham Bioinformatics - FastQC A Quality Control tool for High Throughput Sequence Data'. Accessed: Jun. 03, 2024. [Online]. Available: https://www.bioinformatics.babraham.ac.uk/projects/fastqc/
[62] T. K.Yu *et al.*, 'A bioinformatic pipeline for NGS data analysis and mutation calling in human solid tumors', *Biomed. khimiya*, vol. 63, no. 5, pp. 413–417, Oct. 2017, doi: 10.18097/PBMC20176305413.
[63] D. S. K. Nayak, J. Das, and T. Swarnkar, 'Quality Control Pipeline for Next Generation Sequencing Data Analysis', in *Intelligent and Cloud Computing*, D. Mishra, R. Buyya, P. Mohapatra, and S. Patnaik, Eds., Singapore: Springer Nature, 2022, pp. 215–225. doi: 10.1007/978-981-16-9873-6_20.
[64] P. D. John Marshall, 'Samtools - Duplicate Marking', SAMtools. Accessed: Jun. 17, 2024. [Online]. Available: https://www.htslib.org/algorithms/duplicate.html
[65] X. He *et al.*, 'Comprehensive fundamental somatic variant calling and quality management strategies for human cancer genomes', *Briefings in Bioinformatics*, vol. 22, no. 3, p. bbaa083, May 2021, doi: 10.1093/bib/bbaa083.
[66] D. Mapleson, G. Garcia Accinelli, G. Kettleborough, J. Wright, and B. J. Clavijo, 'KAT: a K-mer analysis toolkit to quality control NGS datasets and genome assemblies', *Bioinformatics*, vol. 33, no. 4, pp. 574–576, Feb. 2017, doi: 10.1093/bioinformatics/btw663.
[67] A. McKenna *et al.*, 'The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data', *Genome Res.*, vol. 20, no. 9, pp. 1297–1303, Jan. 2010, doi: 10.1101/gr.107524.110.
[68] R. K. Patel and M. Jain, 'NGS QC Toolkit: A Toolkit for Quality Control of Next Generation Sequencing Data', *PLOS ONE*, vol. 7, no. 2, p. e30619, Feb. 2012, doi: 10.1371/journal.pone.0030619.

Author Profile

Monica Rawat is an aspiring biotechnology student currently pursuing Masters of Science in Biotechnology at Amity Institute of Biotechnology Noida. She is actively engaged in a pivotal research project entitled "Assessment and Application of Quality Control Strategies in Next-Generation Sequencing Data for Pediatric Cancer". Monica aspires to further her impact on the field by pursuing a PhD.

Ajoykumar Thongbam is an MTech. Biotechnology student at Amity Institute of Biotechnology, Amity Noida. He is an introspective student deeply passionate about computational and mathematical biology. His academic pursuit is focused on innovating in these disciplines to unearth novel insights into evasive diseases, with the goal of contributing to groundbreaking therapeutic discoveries.

Dr. Rajshree Das is an Associate Professor in Amity Institute of Biotechnology, Amity University Noida. Her research interest is on the molecular analysis of cag pathogenicity island of Helicobacter pylori strains. Her present research is focused to understand the genome and genetics of Gastric microbial community associated with or without Helicobacter pylori in causing Gastrointestinal diseases.