From Samples to Germline and Somatic Sequence Variation: A Focus on Next-Generation Sequencing in Melanoma Research

Adrián Muñoz-Barrera [1,†] , Luis A. Rubio-Rodríguez [1,†] , Ana Díaz-de Usera [1,2] , David Jáspez [1] , José M. Lorenzo-Salazar [1] , Rafaela González-Montelongo [1] , Víctor García-Olivares [1] and Carlos Flores [1,2,3,4,*]

1. Genomics Division, Instituto Tecnológico y de Energías Renovables (ITER), 38600 Santa Cruz de Tenerife, Spain
2. Research Unit, Hospital Universitario Nuestra Señora de Candelaria, 38010 Santa Cruz de Tenerife, Spain
3. CIBER de Enfermedades Respiratorias, Instituto de Salud Carlos III, 28029 Madrid, Spain
4. Facultad de Ciencias de la Salud, Universidad Fernando de Pessoa Canarias, 35450 Las Palmas de Gran Canaria, Spain
* Correspondence: cflores@ull.edu.es
† These authors contributed equally to this work.

Abstract: Next-generation sequencing (NGS) applications have flourished in the last decade, permitting the identification of cancer driver genes and profoundly expanding the possibilities of genomic studies of cancer, including melanoma. Here we aimed to present a technical review across many of the methodological approaches brought by the use of NGS applications with a focus on assessing germline and somatic sequence variation. We provide cautionary notes and discuss key technical details involved in library preparation, the most common problems with the samples, and guidance to circumvent them. We also provide an overview of the sequence-based methods for cancer genomics, exposing the pros and cons of targeted sequencing vs. exome or whole-genome sequencing (WGS), the fundamentals of the most common commercial platforms, and a comparison of throughputs and key applications. Details of the steps and the main software involved in the bioinformatics processing of the sequencing results, from preprocessing to variant prioritization and filtering, are also provided in the context of the full spectrum of genetic variation (SNVs, indels, CNVs, structural variation, and gene fusions). Finally, we put the emphasis on selected bioinformatic pipelines behind (a) short-read WGS identification of small germline and somatic variants, (b) detection of gene fusions from transcriptomes, and (c) de novo assembly of genomes from long-read WGS data. Overall, we provide comprehensive guidance across the main methodological procedures involved in obtaining sequencing results for the most common short- and long-read NGS platforms, highlighting key applications in melanoma research.

1. Introduction

Cutaneous melanoma is the major culprit in skin cancer-related mortality, as it is a highly aggressive skin tumor with the highest mutation load among tumors [1]. As with any other type of cancer, cutaneous melanoma could have a somatic (i.e., sporadic) or a germinal (i.e., familial) origin. The first one is the most common form, explaining ~90% of all melanoma cases, and it is caused by weak- or moderate-risk somatic mutations [2,3]. These could explain why sporadic cutaneous melanoma shows a clear relationship with risk factors such as the presence of naevi [4–6], exposure to UV irradiation [7–10], and with polygenic factors such as fair skin [11,12], among others. In addition, the most frequent and well-known genetic alterations occurring in melanoma are linked to the *BRAF*, *NRAS*, *KIT*, and *NF1* genes [13–16]. The familial form of cutaneous melanoma (i.e., in families with

at least another relative affected) has an incidence of ~8% of the cases. In this respect, the vast majority of the highly penetrant germline mutations mainly affect the *CDKN2A* and *CDK4* genes [17].

The advent and adoption of next-generation sequencing (NGS) technologies have accelerated the development of human genomics and personalized medicine, allowing us to study the role of both germline and somatic mutations more precisely in disease. This has facilitated the increase in knowledge of most cancer types, including melanoma, through genomic, transcriptomic, and epigenomic approaches. The decrease in costs and the increase in coverage of targeted gene panels, whole-exome (WES), whole-genome (WGS), and transcriptome (RNA-Seq) sequencing applications offer the possibility of rapidly improving clinical studies, triggering novel and more comprehensive analyses in cancer research [18]. Recent advances in long-read sequencing or third-generation sequencing (TGS), such as those provided by Pacific Biosciences (PacBio) and Oxford Nanopore Technologies (ONT), are nowadays booming because of the facilitation of studies of somatic mutations that affect large and complex regions of the genome that would be difficult to analyze otherwise with the more standard short-read sequencing approaches [19,20].

This review provides a summary of the technical details involved in library preparation for the sequencing process, describing different alternatives for assessing tumor tissues, as well as outlining the NGS and TGS technologies and their use in the study of germline and somatic variation in cancer, with a focus on melanoma. For that, we have performed a systematic review of the most recent literature in search of studies that have applied high-throughput sequencing methods in cancer studies, especially those focused on cutaneous melanoma. Only studies in humans and written in English were included in this assessment. The search was performed in the NCBI PubMed from May to October 2022, using the following terms: "melanoma", "cutaneous melanoma", "cancer genomics", "DNA library preparation", "RNA library preparation", "next-generation sequencing melanoma", "somatic bioinformatics", "somatic pipelines", "somatic structural variation", "structural variation in cancer", "NGS quality control", "kinship estimation", "transcriptomics genefusion", "long reads cancer genomics". Besides this, and given that we aimed to review many steps of the methodology, we also revised articles that were published after 2010 until October 2022. For obvious reasons, this excludes the specific literature reviewed that focuses on the classical sequencing approaches or previous technical methods. A detailed description of the steps involved in selected bioinformatic workflows for short and structural variant discovery is also provided, both for short and long-read technologies, highlighting scripting languages and pipeline editors considered standards in the field and detailing the most commonly used tools and databases that are needed to functionally annotate and classify the discovered variants.

2. DNA Libraries

In the context of cutaneous melanoma, sequencing techniques used have included, among others, targeted sequencing (focusing on specific regions of the genome when prior information is available), WES (limited to the gene-coding regions and alike), and WGS (to detect alterations in the coding and non-coding regions of the genome). Studies leveraging WES and WGS have helped to identify genes that are important for melanoma pathogenesis and, for example, for improving the classification of the different molecular subtypes of melanoma [21].

The typical NGS workflow comprises different steps, from nucleic acids extraction to variant annotation. The process generally begins with converting nucleic acids (RNA or DNA) from biological samples to a biomaterial compatible with the sequencing system intended for the study. This first step is referred to as library preparation—a library is a set of DNA fragments with attached adapters—and it is one of the most important steps, having key biological and bioinformatics implications [22]. Some of the main factors to consider in obtaining high-quality sequencing libraries are the quantity and integrity of the starting material, and the application to be performed. One of the difficulties when working

with melanoma-affected tissues is that the starting material is degraded or of a limited amount. Most excised melanoma lesions are small, from 1 to 2 mm in thickness, and the entire tumor requires formalin-fixed and paraffin wax embedded (FFPE) for diagnosis by histopathologic examination, typically precluding the availability of the frozen tissue, which is more optimal for research. DNA and RNA extraction methods for FFPE tissues vary in the quality and quantity of the resulting material, which may impact the performance of downstream assays. Fixation protocols vary between laboratories. The type of fixative, temperature, pH, chemical crosslinking, or exposure time to formalin and how they are handled contribute to potential nucleic acid damage [23,24].

DNA from FFPE tumor tissues is fragmented, often in low concentration. In melanoma research, the purified DNA can also be contaminated with the pigment melanin, which inhibits polymerase activity [25]. After fixation, DNA fragmentation also changes and increases over time [26], even under certain storage conditions [27]. Usually, the amount of damage in FFPE tissue correlates with the age of the sample. The use of melanoma FFPE samples in amplicon-based NGS panels has shown that storage time was the most critical variable that influenced sample viability for library construction. In this case, the incorporation of quality control (QC) steps and a measure of the DNA integrity (DIN, DNA Integrity Number) helped refine the rate of conversion from samples to NGS results, and particularly to identify which of the oldest samples could be used in the study [28]. Moreover, formalin-induced deamination can lead to artifactual cytosine (C) to thymine (T) and guanine (G) to adenine (A) (C:G > T:A) mutation calls. The proportion of deaminated C bases by formalin fixation is low, generating false low-frequency single nucleotide variants (SNVs). These low-frequency mutations also occur naturally in the tumor process and may be of clinical importance. Therefore, it is essential to repair the deamination in FFPE DNA samples before continuing the rest of the process [29].

Library preparation methods are of the utmost importance when only a small amount of starting material is available and clinical samples are precious. The starting material is generally isolated double-stranded genomic DNA, and the DNA is enzymatically or physically fragmented, followed by end-repair and adapter ligation. Adaptor ligation is followed by size selection to remove free adapters and select the libraries in the desired size range. PCR amplification could also be performed in the resulting selection to obtain enough template DNA for accurate quantification and to further enrich the libraries. However, the amplification step is known to introduce some bias together with fragmentation and size selection [30]. Alternatively, PCR can also be used to add the adapter sequence using tailed primers, which generate molecules with all the elements necessary for sequencing.

Moreover, obtaining the highest possible level of sequence complexity in an NGS library is crucial, as this will reduce the amount of bias. Library complexity refers to the number of unique DNA fragments that are present, i.e., the library should reflect the starting material as closely as possible. The loss of complexity, derived from using PCR, increases the number of duplicate reads. Moreover, shorter fragments are less specific in the bioinformatic alignment against the genome reference step and, thus, decrease the complexity of a sample. In addition to the above PCR considerations, the presence of melanin could inhibit the reaction [31] by forming reversible complexes with DNA polymerase [25]. Additional treatments that allow the proper use of PCR have been described, such as the addition of bovine serum albumin (BSA), DNA dilutions, and DNA purifications with the NucleoSpin® gDNA Clean-up XS kit [32]. A study supported that centrifugation combined with the OneStep™ PCR Inhibitor Removal Kit (Zymo Research Corp, Irvine, CA, USA) was the best method to obtain adequate material for sequencing [33].

The preparation of a library depends on the sequencing platform and the approach. However, in general terms, among the steps to generate the libraries, the fragmentation methods, the attachment of the adapters, and the quantification and library size determination should be considered. DNA fragmentation can be performed by physical, chemical, or enzymatic methods. Physical fragmentation is usually carried out by sonication, in which high-frequency acoustic energy is focused on the DNA sample to break up the

molecules. In enzymatic fragmentation, the restriction endonucleases are the activities involved in fragmenting the DNA. An alternative enzymatic method for library preparation is tagmentation, which uses the transposase enzymatic activity to fragment DNA while adding specific adapters to both ends of the fragments (Illumina, San Diego, CA, USA). Therefore, it improves traditional preparation processes by combining DNA fragmentation, terminal repair, and adapter ligation in a single step, thus reducing the hands-on time. The attachment of adapters to the ends of the DNA molecules allows the identification of each processed sample. However, the existence of a high proportion of unattached adapters can cause adapter dimer problems. If these are not removed, they may result in a significant reduction in sequencing quality and efficiency. One of the most extended processes for their elimination consists of using magnetic bead-based clean-up steps. Regarding the fragment (also known as an insert) size, the optimal size is determined by the limitations of the NGS instrumentation and the specific sequencing application. With the current Illumina, Inc., technology, the optimal insert size is affected by the cluster generation process, where shorter products are amplified more efficiently than longer products. Therefore, assessing the fragment distribution of the final libraries is an essential QC step to ensure optimal results. This step could be automated using electrophoresis systems such as the TapeStation instrument (Agilent Technologies, Santa Clara, CA, USA). The evaluation of the quality of the final libraries is a critical step. Accurate quantification is essential since it provides an estimate of the molecules available to be sequenced in each sample. Quantification can be carried out using different methods, such as intercalating dyes, hydrolysis probes, droplet digital emulsion PCR, or fluorometry.

3. RNA Libraries

While this review has a focus on detecting and studying somatic and germline variation in melanoma, specific applications for assessing large structural variations that are important for cancer research could be based on transcriptomics, and some are discussed in Section 5.5. Because of that, we also provide some basics in case the starting material is RNA. For transcriptomic studies based on RNA-Seq, the typical steps include isolating the desired RNA molecules, reverse transcription to complementary DNA (cDNA), fragmentation or amplification of randomly primed cDNA molecules, and ligation of sequencing adapters [34]. The accuracy of gene expression quantification depends on the purity of the samples, and tumor tissue samples often comprise disease-state cells surrounded by normal cells. RNA library preparation also requires high-quality RNA isolated from the biological sample. RNA quality is commonly measured with a bioanalyzer (Agilent Technologies) or a TapeStation system, which provides an RNA Integrity Number (RIN) between 1 and 10, with 10 being the highest quality with minor degradation. Low RNA quality (RIN < 6) can strongly affect the sequencing results [35]. Alternatively, the quality of isolated RNA can be evaluated qualitatively based on the presence of intact ribosomal RNA (rRNA) bands on an agarose gel. As previously mentioned, FFPE tissues are usually of poor quality. Thus, the effect of RNA degradation must be carefully considered in the sequencing results [36]. Several commercially available solutions are well suited for FFPE and low-quality input samples.

The next step in RNA-Seq is library creation, which starts with the removal of remanent DNA and the isolation of the desired RNA molecules. There are several options in RNA-Seq library construction and experimental design to fit the specific needs of the researcher (poly-A selection, ribo-depletion, size selection, strand-specific, duplex-specific nuclease, multiplexed, short or long reads). In general, most RNA molecules in tissues are rRNA. Therefore, to detect less-abundant RNAs and for cost efficiency, it is necessary to remove rRNA transcripts before library construction. This rRNA depletion step avoids the consumption of the sequencing reads by rRNAs, increasing the overall depth of coverage of the RNAs of interest. Alternatively, messenger RNAs (mRNAs) are enriched by selection for polyadenylated (poly-A) RNA. The 3′ poly-A tail of mRNA molecules is targeted using poly-T oligos covalently attached to magnetic beads. Each methodological approach

presents technical biases and limitations. Poly-A libraries are the best option for obtaining the coding RNA transcripts. While using rRNA depletion helps to accurately quantify non-coding RNAs and post-transcriptionally unmodified pre-mRNAs. Moreover, there are specific protocols for selectively targeting small RNA species, which are key regulators of gene expression. Small RNA species (15–30 nucleotides) lack poly-A and are microRNAs (miRNAs)—more than 800 miRNAs are deregulated in melanoma [37]—small interfering RNAs (siRNAs), and Piwi-interacting RNAs (piRNAs). Isolated RNA, with high quality and enough amount, is then fragmented, randomly primed, and subjected to the first and second cDNA strand synthesis. Finally, the adapters are ligated to the ends of cDNA fragments and amplified.

As in the case of DNA libraries, the protocol must add some QC steps. One consists of verifying the library profiles in an Agilent TapeStation system to ensure that their size is in the appropriate range and to determine the presence of unexpected peaks. The other QC is quantification, such as by quantitative PCR (qPCR), fluorometry, or the Agilent TapeStation system. The best method is qPCR since it quantifies the complete libraries, that is, those that can form clusters in the sequencing flow cell.

4. Sequencing-Based Approaches in Cancer and Cutaneous Melanoma Research

4.1. Sequencing with the Classic Approaches

The first generation, or Sanger sequencing [38], has been used to detect disease-causing variants [39–41], allowing the assessment of DNA fragments up to 1000 base pairs (bp) [42]. Being considered the gold standard in clinical research and key for assembling the first draft of the human genome for the Human Genome Project [43], the Sanger method has been used in melanoma research to characterize, for example, the particular behavior of it in different populations such as Taiwanese [44] and Chinese [45], among others. Some studies have shown a comparable performance between the first-generation sequencing with other variant detection approaches [46], with even some gains in efficiency and sensitivity in the case of the Sanger method [47].

In spite of the accuracy of the first-generation sequencing and the capability to evaluate repeated elements [48–50], the impressive development and throughput improvement in NGS have pushed aside the use of Sanger sequencing [42]. In this regard, pyrosequencing, the forerunner of NGS approaches, employs luminescence to identify nucleotides of the DNA strand based on the sequencing-by-synthesis (SBS) principle [51]. This sequencing technology has been used in melanoma research to unravel the clinical phenotypes related to *NRAS* and *BRAF* mutations [52,53]. Different commercial protocols have been developed to identify the most common mutations in codons within the *BRAF* gene, such as Therascreen™ BRAF Pyro Kit (Qiagen Inc., Valencia, CA, USA) for mutations in codons 464, 469, and 600. Additional molecular protocols for codons in *BRAF* ("BRAF Codon 600 Mutation Detection by Pyrosequencing"), *KRAS* ("KRAS Mutation Detection"), and *NRAS* ("NRAS Mutation Detection by Pyrosequencing") have been conceived by ARUP Laboratories (Salt Lake City, UT, USA) to help in the treatment of patients with different solid tumors, including melanoma.

4.2. Next-Generation Sequencing

In 2005, a new sequencing system was released to the market based on pyrosequencing and emulsion PCR, allowing the parallelization of amplification reactions for the first time and a quantum leap in scale at the performance level [54]. This emerging technology, considered the first NGS system, opened the horizon for the development of many other approaches, which ultimately have resulted in an array of applications in clinical practice and biomedical research [55–57].

Since then, several others have emerged in this decade. Based on the sequencing chemistry, one can distinguish between sequencing-by-ligation (SBL), which uses a DNA ligase to add the nucleotides to the newly synthesized DNA molecule [58], and SBS, which uses a DNA polymerase instead of a ligase [59] (Table 1). An example of the first type is

the SOLiD technology (Sequencing by Oligonucleotide Ligation and Detection) (Thermo Fisher Scientific, Waltham, MA, USA), whereas there are various commercial SBS-based sequencers, including Ion Torrent (Thermo Fisher Scientific, Waltham, MA, USA), MGI Tech (Shenzhen, China), or Illumina, Inc. (San Diego, CA, USA), among others. This review will focus on the applications based on the latter, since it is dominant in the market, because of its high versatility, performance, and market competitiveness. Despite this, it is worth mentioning that MGI Tech has become more and more popular in recent years because of its reduced costs and increased performance.

Table 1. Overview of the most common commercial NGS platforms and instruments.

Brand	Instrument	Key Applications	Run Time (h)	Max. Output (Gb)	Max. Read Length (Bases)
Illumina, Inc.	NextSeq 550	Targeted Gene Sequencing Transcriptome Sequencing	12–30	120	PE150
	NextSeq 1000 and 2000	WGS (limited samples) WES Targeted Gene Sequencing Transcriptome Sequencing	11–48	360	PE150
	NovaSeq 6000	WGS WES Targeted Gene Sequencing Transcriptome Sequencing Methylation Sequencing	13–44 *	6000	PE250
	NovaSeq X Series	WGS (large sample number) WES Targeted Gene Sequencing Transcriptome Sequencing Methylation Sequencing	13–48 *	16,000	PE150
MGI Tech	DNBSEQ-G50	Targeted Gene Sequencing	9–40	150	PE150
	DNBSEQ-G400	WGS (limited samples) WES Transcriptome Sequencing	13–109 *	1440	PE300
	DNBSEQ-T7	WGS (large sample number) WES Targeted Gene Sequencing Transcriptome Sequencing	24–30	6000	PE150
Ion Torrent	Ion GeneStudio S5/Plus/Prime	WES (limited samples) Targeted Gene Sequencing	6–19	15/30/50	SE200/SE400/SE200
	Genexus System	WES (limited samples) Targeted Gene Sequencing	2–3	15	SE200

* Depends on the flow cell used. WES, whole-exome sequencing; WGS, whole-genome sequencing; Gb, gigabases; PE, paired-end; SE, single-end.

NGS allows to read billions of base pairs of DNA sequences quickly and simultaneously in only one experiment ("run"), resulting in a large dataset and an important cost efficiency. Routinely, an Illumina NGS experiment can be divided into the following four main steps: (1) fragmentation, (2) indexing or attachment of the adapters, (3) amplification, and (4) sequencing. As the two first steps have been extensively explained in the library preparation sections, we will now focus on the amplification and sequencing steps. One of the features of short-read approaches is the need for a PCR step prior to the sequencing run, which allows to establish the different clusters where the sequencing will take place. In Illumina instruments, classical amplification is produced by a bridge PCR. It means that one of the extremes of the single-stranded library attaches, by sequence complementarity, to one of the multiple single-stranded oligonucleotides on the coated surface of the flow cell. As this hybridization occurs, amplification begins immediately. A double-stranded molecule

is provided, which is denatured, followed by a washout of the original template, whereas the covalent attachment of the newly synthesized strand is kept. This new molecule flips over and creates a bridge with a complementary oligonucleotide from the surface of the flow cell. Once a single-stranded library is bridged, the amplification starts and the library becomes double-stranded, coining this as the bridge amplification. Next, the denaturation step renders two single strands covalently bound to the flow cell. The bridge amplification continues until all oligonucleotides have been used. Afterward, linearization is carried out, the reverse strands are cleaved and washed away, and the forward strands are maintained on the surface. Finally, the 3′ extremes of the amplified fragments are blocked, and the sequencing primer is added to start the sequencing process. In the high-end of the through-put scale of Illumina (i.e., HiSeq 4000 and NovaSeq 6000 sequencing platforms currently) this process is also an exclusion amplification (ExAmp) to ensure that only one molecule attaching to each of the flow cell microwell forms a cluster. The patterned flow cells are also one of the exclusive features of the HiSeq 4000 and NovaSeq 6000 systems.

The NGS technology has been widely applied in cancer genomics, most commonly using short-read technologies and a high depth of coverage to study somatic variation. Based on that, the analysis of cancer samples, including melanoma subtypes [60–62], is typically performed using targeted sequencing of a cancer-specific gene panel, WES, or WGS (Figure 1).
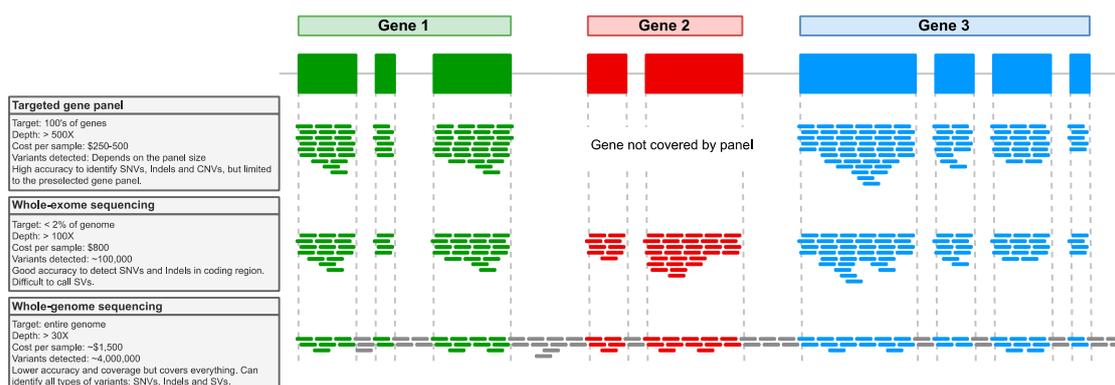
Figure 1. Overview of NGS DNA sequencing-based experiments applied on cancer genomics.

Panel sequencing reduces costs, enables faster turnaround times, and requires a less complex pipeline for variant detection. Due to the remarkable benefits of this approach, different laboratories or manufacturers have developed their own panels to solve specific questions related to many types of cancers, such as the Hereditary Cancer Solutions by SOPHiA GENETICS (Boston, MA, USA), which allows assessing breast and ovarian cancer and some others involved in cancer-associated or predisposition genes in gastric [63] or pediatric cancer [64], among others. However, it has the disadvantage of not allowing the analysis of genes or genomic regions originally not included in the panel. Typically, these panels only cover driver mutations in genes known to be involved in melanoma, such as *BRAF*, *NRAS*, *KRAS*, *KIT*, *GNAQ*, and *GNA11* [65], and do not include genes that have been recently found by WES/WGS studies [16,66].

In a recent germline-focused study, the authors used WES and targeted gene panel sequencing of uveal melanoma samples, identifying associated susceptibility genes, and suggesting a locus heterogeneity in hereditary predisposition [62]. In another study, new therapeutic targets potentially related to alternative splicing caused by somatic mutations in multiple genes were specifically identified through WES [60]. Likewise, Vergara et al., using both WGS and WES data, analyzed the evolution of human melanoma from early to late-stage disease and found that it was dominated by tetraploidization and large-scale acquisition of aneuploidy [61].

WES enables assessing the mutational spectrum from virtually all the protein-coding regions, which harbor ~2% of the genome [67]. This cost-effective application allows the

analysis of SNVs and small insertion-deletion variants (indels, around <50 bp in size) with high coverage reads. This implies that the data is more manageable, although of limited use to cover and identify larger structural variations (SVs) [68,69] that have key implications in melanoma [70,71]. Furthermore, familial pancreatic cancer [72], recurrent prostate cancer [73], malignant ovarian germ cell tumors [74], familial colorectal cancer [75], and locally recurrent rectal cancer [76] are just some of the cancers where the implementation of WES offers benefits by allowing the discovery of putative predictors and identifying risk genes and potential driver mutations involved in the pathogenesis. Whereas WGS allows to cover virtually all variation across the genome and to better unveil SVs. To date, several WGS studies have been carried out, focusing on the analysis of SVs in different subtypes of melanoma [61,66,77,78], but also to assess how mitochondrial genetic variation could influence gastric cancer [79], the discovery of novel mutations involved in prostate cancer [80], or even to study how the treatment could impact metastatic colorectal cancer [81]. In particular, the studies in melanoma have been able to demonstrate and identify key non-coding regions that are involved in the progression or risk of the disease and that cannot be detected with gene panels or WES.

5. Bioinformatic Workflows for NGS Data Analysis

Different advances and approaches in NGS are common in the toolbox of studies in cancer genomics and, specifically, in melanoma research [82–85]. NGS can generate different types of data, allowing to detect a wide variety of genomic abnormalities simultaneously. Besides, this technology can also help to analyze the molecular mechanisms of cancer, identify somatic mutations that have accumulated during tumorigenesis, and even assist in the discovery of new genomic, transcriptomic, and epigenomic profiles of individual malignant growths [82,83]. Likewise, the treatment of cancer patients could be improved based on NGS data, constituting one of the pillars of precision oncology [86,87].

In an effort to characterize cancer genomic alterations and their diversity, initiatives such as The Cancer Genome Atlas (TCGA) [88,89] and the International Cancer Genome Consortium (ICGC) [90] have gathered a vast number of cancer genomes from patient samples around the globe. In addition, the Genomics England 100,000 Genomes Project Cancer Program was established to develop a national molecular data research platform linked to longitudinal clinical data and to transform the National Health Service's clinical cancer care based on WGS data [91,92]. Information collected as part of these studies includes clinical data, raw genomic data, and processed data. These data have been used not only for characterizing the mutational landscape of melanoma [93–95], but also to reveal mechanisms of tumor spreading [96], and identify biomarkers of treatment response [97], among many others.

A common method is to analyze paired normal and cancer tissues from the same patient and use the normal as a comparator [98]. As sequencing costs continue to decline, sequencing platforms are being redesigned to prioritize WGS variant reporting based on clinical relevance [99]. Thus, current trends are focusing on developing new bioinformatic algorithms for both WES and WGS in order to improve their clinical application [18,100]. Interestingly, developments and improvements in algorithms in the context of cancer genomic data analysis make it possible to provide a probability score of disease-driving mutations and to identify other potential targets [101,102].

Given the large amount of data that is generated as part of the NGS experiments, managing, administering, and storing such large sequence datasets, as well as the need for an efficient analysis, is a real challenge. Storage requirements for raw, intermediate, and processed data critically depend on the type of experiment as well as a number of parameters, such as the depth of coverage or the number of different variant detection tools that will be involved. As expected, experiments based on tumor-normal pairs or on WGS need a larger amount of disk space. Bioinformatic analysis typically begins with the raw data sequencing and ends with a listing of somatic variants per sample or aggregate across the samples of the experiment. These steps include processing raw reads, alignment to

the reference genome, variant calling, annotation, filtering, and prioritization of variants. However, the software tools and workflows to be used depend on the type of experiment and the processing strategy applied. Currently, there is no single gold standard processing strategy for cancer data, and each pipeline implements these steps, or most of them, using different tools and parameters. Some of the state-of-the-art workflows for cancer genomics are the NYGC Cancer Pipeline [103], Sarek [104], and the GATK Best Practices for both somatic short and copy number variant (CNVs) discovery [105].

Moreover, the process of aligning reads with the reference human genome, variant calling, and assembly for cancer WGS data requires a large amount of computational power at each analysis step (Figure 2) [106]. In this context, for faster data throughput as well as to significantly simplify and reduce disk space requirements, it is key to focus on combining as many of these steps or computational tools as possible using Unix pipes. In this way, for workflow standardization and automation, several managers are available, such as Snakemake [107], WDL [108], or Nextflow [109]. To overcome the limitations of hardware and support required for large-scale genomics projects, there are high-performance computing (HPC) facilities. These are equipped with a cluster of high-speed computing nodes and multi-petabyte storage systems, enabling distributed and parallel computing, cloud computing, and graphics processing unit (GPU) computing, among others.
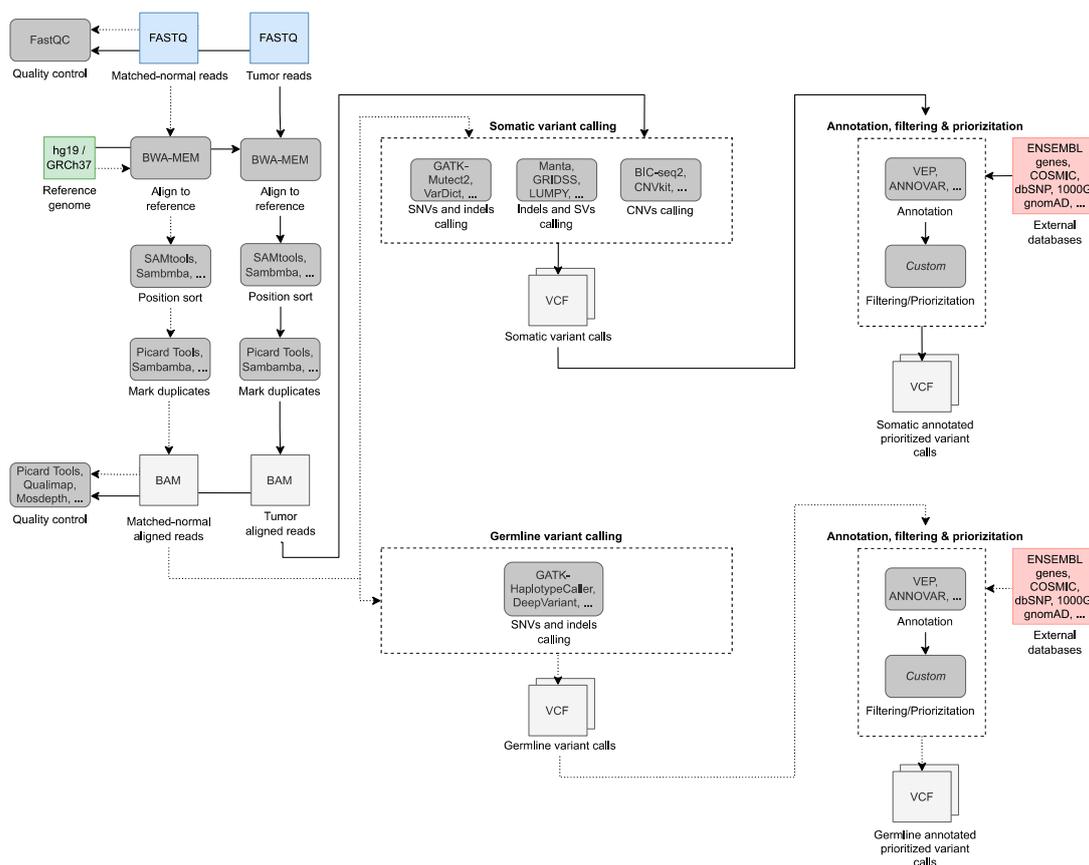
Figure 2. Schematic overview of the pipeline for WGS somatic mutation identification.

From here on, each step of a typical pipeline for cancer genomics will be briefly described, indicating the most common tools and peculiarities for different types of experiments.

5.1. Read Alignment to the Reference Genome

Paired-end read sequence data is generally provided as two files in FASTQ format, each file representing one end of the read. The sequence data is stored in the FASTQ files as plain text and contains the sequence of the read and the per-base quality scores.

In a typical pipeline, the sequence files are aligned to the reference sequence using an aligner. Different builds of the human reference genome are available, with GRCh37 (hg19) and GRCh38 (hg38) being the most popular. In practice, hg19 is still the most widely adopted, as most vendors provide their probesets in hg19 coordinates [110]. Recently, the Telomere-to-Telomere Consortium (T2T) has generated a gapless reference of the human reference genome [111]. Besides, the T2T-CHM13 assembly corrects and expands the sequence coverage of the GRCh38 human genome by more than 200 Mbp, including highly repetitive DNA sequences at telomeres and centromeres of the 22 autosomes and the X and Y chromosomes [111]. This new assembly has enabled many previously unknown genes to be identified and has been shown to significantly reduce false positives in hundreds of medically relevant genes [112]. A wider adoption of this new assembly, including in clinical practice, will depend on creating new annotations and a liftover of the existing major genome annotations to T2T-CHM13 [113].

To perform short-read alignment of gene panel, WES, or WGS data, BWA-MEM [114] is one of the most widely used aligners. Other popular alignment tools include Novoalign [115], Bowtie2 [116], or Minimap2 [117]. The resulting aligned sequences and their related metadata are stored in the SAM/BAM file format (Sequence Alignment Mapping). This file is subsequently sorted by genomic coordinates and indexed for quick access. SAMtools [118] is the most used tool to manage SAM/BAM files since it allows us to carry out most of the operations. Marking or removing read duplicates in the BAM file is a crucial step to account for PCR duplicates of the exact same DNA fragment and limit their impact in the variant calling stage. Tools such as Sambamba [119] and Picard [120] are commonly used to identify and mark duplicate reads in SAM/BAM files to exclude them from subsequent analyses. Downstream analyses rely on the SAM/BAM files to identify a wide range of genetic variations. QC steps of SAM/BAM files should be made prior to variant calling to evaluate the sequencing metrics, assess the depth of coverage and the percentage of duplicate reads, evaluate sample contamination, or perform sex inference.

Depth of coverage is a key metric to evaluate, often defined as the average number of non-duplicated reads that align across the target region. The target region can be the exons of a gene panel, the targeted exons across all genes, or even the entire genome. The level of coverage often determines whether variant discovery can be performed with a certain degree of confidence at a given genomic position. Typically, the coverage needed for an NGS experiment is determined by the method being used and the characteristics of the experiment. This metric should be calculated on both normal and tumor BAM files and can be easily obtained using tools such as Mosdepth [121]. For gene panel and WES data, there are also a number of key parameters to consider, as indicated elsewhere [67]. On-target mapped reads and on-target coverage should be calculated to assess potential problems during the library preparation. Picard Tools and Qualimap [122] are usually used for this purpose. MultiQC [123] can also be helpful to aggregate QC results from different bioinformatic tools and different experiments into a single report.

5.2. Variant Calling of SNVs and Indels

The alignment results in BAM format are subsequently examined for the presence of any type of somatic variation. The accurate identification of mutations is of critical importance. Numerous variant callers are available for this purpose. A list of the most widely used SNV and indel callers can be found in Table 2. To distinguish germline from somatic mutations in the tumor, a common practice is to rely on a normal tissue sample from the same individual. Somatic callers such as GATK-Mutect2 [124], Strelka2 [125], and VarScan2 [126] consider simultaneously the aligned data from the tumor and normal samples.

Sequencing the tumor and normal sample genomes allows not only the identification of variants with greater fidelity but also allow the finding the of potential therapies (see references in Table 2). If sequencing data from the normally matched sample is available, it is also recommended to run a germline variant calling to detect variants that may

indicate possible susceptibility to cancer or may be useful in treatment responses. GATK HaplotypeCaller [105] and DeepVariant [127] are widely used tools for this purpose.

Table 2. Most common tools to call somatic variants and related research studies in cancer genomics.

Somatic Callers	Sequencing Approach			Type Mutations	Normal Sample Required in Somatic Mode	Related Somatic Studies
	Targeted	**WES**	**WGS**			
GATK-Mutect2 [105]	✓	✓	✓	SNVs and indels	Optional	Liver cancer [128], lung cancer [129]
Strelka2 [125]	✓	✓	✓	SNVs and indels	Yes	Cervical cancer [130]
VarDict [131]	✓	✓	✓	SNVs and indels	Optional	Breast and ovarian cancer [132]
CNVKit [133]	✓	✓	✓	CNVs	No	Melanoma [134]
Manta [135]	✓	✓	✓	SNVs and indels	Optional	Gastric cancer [136]
Delly [137]	x	x	✓	SVs	Yes	Plantar melanoma [138]
Lumpy [139]	x	x	✓	SVs	Optional	Colon cancer [140]
GRIDSS [141]	✓	✓	✓	SVs	Yes	Myeloid leukemia [142]
Varscan2 [126]	x	✓	x	SNVs and indels	Yes	Uveal melanoma [143]
ClinCNV [144]	✓	✓	✓	CNVs	Yes	Cutaneous leukemia [145]
ExomeDepth [146]	✓	✓	x	CNVs	No	Breast cancer [147]
ClinSV [148]	x	x	✓	SVs	No	Breast cancer [149]

WES, whole-exome sequencing; WGS, whole-genome sequencing; SNVs, small nucleotide variants; indels, insertion-deletion variants; CNVs, copy number variants; SVs, structural variants.

As an important remark, some somatic variant calling tools require the normal matched sample, conditioning the choice of the variant caller. When the normal matched sample is not available or useful (i.e., due to technical reasons), several tools allow the use of a "panel of normals" (PoN), made out of sequencing data from normal unrelated individuals (N ~50). The PoN can be used to filter out variant calls associated with recurrent technical artifacts, systematic noisy positions, and germline variants, although its effect is limited since this approach does not eliminate the germline variants of the individual.

The performance of somatic variation detectors varies widely, as demonstrated in various benchmarking studies, each showing strengths and weaknesses [150,151]. The precision of the detection depends mainly on the sequencing depth in each genomic region and on the alignment or mapping error. Considering the complexity of the human genome, especially in non-coding regions, mapping short reads to repetitive regions and tandem repeats typically imposes difficulties, resulting in reduced sensitivity and specificity of most variant detection tools. Because no somatic variation detector has yet surfaced as the gold standard because of a superior performance across all scenarios, a joint approach that combines the results of two or more complementary callers provides a better balance between sensitivity and specificity [150].

Indels have an important role in tumorigenesis, especially if they affect the coding region, where they can significantly disrupt the reading frame and lead to changes in protein function. Because indels have not been studied as thoroughly as SNVs, tools, and methods for indel detection typically need to be fine-tuned and optimized. In this context, initiatives such as NCTR Indel Calling from Oncopanel Sequencing Challenge (https://precision.fda.gov/challenges/22, accessed on 3 September 2022) aim to improve indel detection by validating and benchmarking indel calling pipelines across laboratories.

After the variant calling step, the resulting variant callset is typically reported in variant call format (VCF), encoding metadata and variant records for each sample. VCF files are often compressed and indexed so that they take up less disk space and can be

handled more efficiently by applications. Widely used tools for managing VCF files are BCFtools [118] and VCFtools [152].

5.3. Variant Calling of SVs and CNVs

Structural alterations, including large insertions and deletions, duplications, inversions, translocations of at least 50 bp in size, and gene fusions, have been associated with cancer pathogenesis. Large deletions and amplifications, occasionally spanning genes, or even entire chromosomes, sometimes lead to alterations in gene copy number. This type of SV is usually referred to as CNVs and copy number aberrations (CNA). In most cancer types, including melanoma, a remarkable number of somatic CNAs accumulate during the progression of the disease and have been associated with cancer prognosis and development. CNAs have been directly associated with the expression of driver genes, where copy number changes may increase the expression of oncogenes and decrease the expression of tumor suppressor genes [153–155].

For SV calling based on NGS paired-reads data, SV detection tools typically rely on one or a combination of the following approaches: (a) coverage depth (RD), in which changes in coverage may imply an SV, (b) discordant read pairs (RP) in the alignment, where read pairs map at unexpected distances or orientations, (c) split-read mapping (SR), in which part of the read aligns to either side of an SV, (d) and the assembly approach (AS), which detects SVs by assembly-based sequence reconstruction. The best-performing detection tools usually leverage a combination of some of the above methods [156].

Popular tools for SV detection include Manta [135], DELLY [137], LUMPY [139], GRIDSS [141], and CNVKit [133] for WGS data (see Table 2), some of them being adaptations from germline CNV/SV calling. Many of these tools have been benchmarked and ranked in review studies, providing mixed conclusions [157,158].

In targeted sequencing of a gene panel and WES, only the approach based on the variation in depth of coverage (RD) can be applied; hence, only CNVs can be reliably detected in such experiments. The sparse distribution and small size of exon targets make the relationship between copy number and depth of coverage more complex, making CNV detection less successful. This discontinuous data is skewed by technical limitations arising from GC content bias, non-uniform sequencing depth, and PCR amplification artifacts. In order to mitigate some of these limitations, CNV callers for target sequencing/WES usually perform multi-sample normalization and implement several model-based approaches by using samples sequenced on the same equipment and with the same sequencing kit for better results. Some of the most used CNV detection tools in these experiments are ClinCNV [144], CNVKit [133], and ExomeDepth [146].

Similar to somatic SNV and indel calling, combining the results of at least two of the tools based on different approaches results in an optimal strategy for somatic SV/CNV calling [159]. In the detection of somatic SVs, a matched normal sample is also usually required to be used as a comparator.

Although the detection of SNVs and indels based on NGS data can be considered routine, the detection of SVs even with WGS data still poses many challenges. This is mainly because a large fraction of SVs are found in difficult-to-map regions of the genome, such as repetitive regions or tandem duplications, which impose uncertainty during the aligning process. Additionally, short reads are often insufficient to resolve complex SVs and long insertions, as these can be smaller than the SV sizes. All these may result in miscalling events or provide false positive and false negative calls. For this reason, linked reads and long-read-based sequencing are increasingly being applied to the detection of SVs to achieve higher levels of sensitivity and specificity in the studies.

This is an active area of interest but is still unresolved. Initiatives such as the precisionFDA challenges are aimed to benchmark the state-of-the-art variant callers in challenging genomic regions, especially those important for medical sequencing [160,161].

5.4. Variant Annotation, Filtering, and Prioritization

After variant calling, the identified variant callset, including SNVs, indels, CNVs, and SVs, needs to be annotated with functional information to assess the biological implications. The accurate identification of somatic variants is essential to provide potential candidates to be used in targeted cancer therapy [162]. This process includes annotation to identify if the variant affects the protein coding sequence of a gene, splicing, and other regions, as well as pathogenicity scoring and effect prediction of being a carrier of the variant. Additionally, the variant callset is typically annotated with existing population information from databases and studies, such as the NCBI dbSNP [163], gnomAD genome/exome [164] frequency data, etc., to identify if a variant has been previously identified by another study of germinal variation or COSMIC [165] to assess if it has been previously associated with any type of cancer. Commonly used variant annotation tools such as Ensembl Variant Effect Predictor (VEP) [166], ANNOVAR [167], or GATK Funcotator [168] annotate variants individually in the VCF file. Depending on the reference transcript used by the tool, typically RefSeq or Ensembl, the functional annotation may vary [169]. Hence, the choice of the annotation tool must be performed carefully.

After the annotation step, the variant calls should be filtered to remove common alignment artifacts and reduce the number of false-positive somatic calls. The annotation information is very useful because it enables the filtering the variant callset. Population filtering is also a common strategy for identifying and filtering likely germline variants from somatic mutation callsets. However, this step must be performed carefully, as common databases such as dbSNP and gnomAD contain several mutations from human tumors, whereas somatic variant catalogs, such as COSMIC, contain germline variants. Similar to SNV and indel calling, CNVs and SVs can be filtered against a PoN to remove variants in highly variable regions and artifacts. Germline SV databases such as gnomAD-SV can also be used to filter SVs that are variable in otherwise healthy human populations.

A manual review of tumor and normal sequencing alignments using visualization tools such as IGV [170] can help in eliminating false positive somatic calls. A manual review of CNVs and SVs can also be performed in the alignment file. This may be useful to resolve ambiguous SV breakpoints, although sometimes the variation occurring is difficult to deduce. SVs with well bioinformatic support are often supported by both discordant read pairs and changes in sequence coverage in specific regions. Tools such as Samplot [171] enable the identification of false positive SV calls using visualizations, whereas Samplot-ML [171] is able to discriminate between true and false deletions using convolutional neural networks (CNN) for image recognition [172].

Besides manual review, a subset of the detected variants can be independently validated by orthogonal approaches, such as Sanger sequencing. Despite the filtering protocols that can be implemented, most NGS methods detect many more candidate variants with likely functional effects than it is possible to validate experimentally as part of a project. Variant prioritization is a common practice to obtain a manageable set of variants. Although variants can be ranked based on various parameters, the prioritization of candidate variants that may be related to the disease is a multifactorial problem, and generally represents a bottleneck in cancer genomics. Open access databases such as the Clinical Interpretation of Variants in Cancer (CIViC) [173] have accumulated and curated information from diverse cancer types and are useful to identify cancer biomarkers and variants that may be used for treatment response. Typically, this task has been performed by experts in the biomedical field. Unsupervised and semi-automated techniques have emerged in recent years [174,175], although none of them has become the gold standard.

5.5. Tumor Clone Identification

High-throughput sequencing enhances the study of tumor evolutionary patterns, enabling the deciphering of all mutations in tumor clones [176]. Tumor clones are clusters of cells that share several somatic mutations, and the evolution of each clone can be represented by the variant allele frequency (VAF) [177]. VAF can be obtained by NGS and

is defined as the percentage of reads that match a specific variant divided by the total coverage at that variant locus [178].

Assuming that almost all somatic variants in tumor cells are heterozygous, the proportion of tumor cells with the same mutation is twice the VAF value. This means that a specific variant present in 80% of tumor cells will have a VAF value of 40%. Using a VAF density plot, tumor clones and the present variants in each clone can be represented by each peak in the plot, helping us to infer and identify clonal progression [179]. Initially, in tumor population cells the VAF of the existent mutations would follow a normal distribution with a value near 50%. This means that a clone cell that carries a mutation set is also present in almost all tumor cells. With the clonal evolution of the tumor cells, new mutations may emerge, causing one of the new clones to carry both the original and the new set of variants that may provide survival advantages to the cell. In this case, if half of the tumor cells belong to the new clone, the VAF density plot would have two peaks at 50% and 25% VAF. With the following tumor progression, some clonal cells obtain a third set of mutations, producing more potentially malignant cells, which can lead to adverse effects such as metastasis. Assuming that a quarter of the tumor cells have the third set of variants, the VAF density plot would have three peaks at 50%, 25%, and 12.5%.

5.6. Gene Fusions

Gene fusions (also known as chimeric transcripts) can be caused by somatic chromosomal rearrangements involving large SVs or chromosomal translocations. Gene fusions have been involved in the progression of a variety of cancers [180], including melanoma [181], lung cancer [182], and breast cancer [183]. These mutations can also be generated at the RNA level by the co-transcription of neighboring genes or by splicing processes from different genes. As such, these rearrangements may be more efficiently associated with NGS-based transcriptomics. In this context, transcriptomic analysis using RNA-Seq data has emerged as a promising solution to identify gene fusions of potential importance for cancer development. A simple computational workflow for accurate detection and characterization of fusion transcripts from RNA-Seq, including alignment and variant analysis steps, is presented in Figure 3.
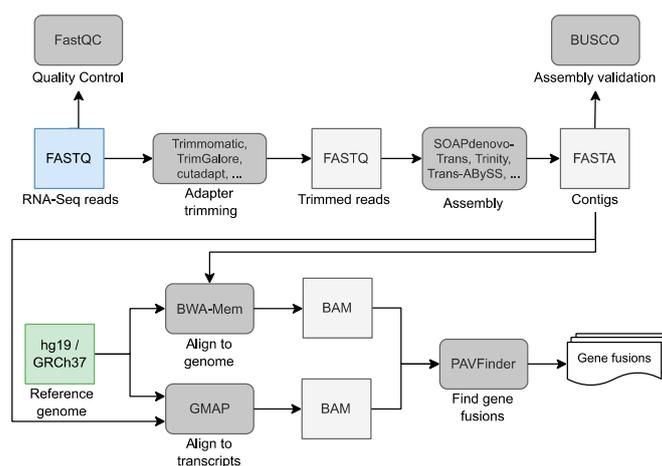
Figure 3. Bioinformatic pipeline for gene fusion discovery through RNA-Seq data. Reads are subjected to a trimming process by means of tools such as Trimmomatic [184], cutadapt [185], or TrimGalore [186] in order to remove adapter sequences in the data. Then, trimmed reads are used for the de novo assembly process, producing assembled contigs in FASTA format. SOAPdenovo-Trans [187], Trinity [188], or Trans-ABySS [189] are examples of RNA-Seq assemblers. These contigs are aligned both to the reference genome and to the reference transcripts. Finally, resulting alignment files in BAM format were analyzed using PAVFinder [190] in order to discover transcriptomic SVs as gene fusions or tandem repeats.

The most common way to shed some light on cancer-related gene fusions is to run de novo assembly and annotation of transcripts using short reads from RNA-Seq [191]. To carry out this process, a variety of transcriptome assemblers [189,192], SV callers, and full pipelines [193] have been developed in recent years. There are also tools such as INTEGRATE [194] that combine WGS and RNA-Seq data from the same sample to discover expressed gene fusions in cancer cells. Methods that are not based on transcriptomes but on de novo assembly of WGS long-read data are also starting to emerge as an alternative way to deeply assess somatic SVs (including gene fusions) in cancer genomes. Further details are provided in Section 6 of this review.

### 5.7. Further Quality Control Steps to Perform in the Callset

The massive amount of data generated by NGS technologies need to establish standardized procedures, guidelines, and rigorous QC steps to ensure accurate results. More importantly, these steps are key for the implementation of NGS technologies in clinical areas, where high-quality data and reliable results are fundamental [195]. Many QC steps are applied throughout the bioinformatic workflow. Different tools, such as FastQC [196] and Qualimap 2 [122], perform the QC based on raw reads and read mapping, respectively. However, these have been covered in previous sections. As important as these are the traceability of samples to verify which results correspond to which samples when processing multiple samples in parallel. For this purpose, studying family relationships within a cohort or inferring sample sex are essential steps to help sample tracking.

5.7.1. Relatedness

Traditionally, pedigrees were the gold standard to infer relatedness, but the advent of new technologies and the availability of different genetic markers make it possible to evaluate relatedness using many approaches. Tools to infer relatedness exist for the SNP array technologies, such as KING [197], REAP [198], or KIND [199], which could be used in this context. Specifically, for NGS data, tools such as Somalier [200] are very convenient to infer family relationships with data from diverse applications (WGS, WES, RNA-Seq, etc.). Relatedness among samples is calculated by allelic concordance from SNVs within these positions (classified as homozygous, heterozygous, and alternative homozygous).

5.7.2. Sex Inference

The genetic inference of the sex of a sample from its sequence obtained in an NGS experiment is a mandatory QC step to detect errors in the metadata, which helps to improve sample traceability. Several bioinformatic tools are available to infer the biological sex of the sample based on the proportion of reads aligning to X and Y chromosomes, such as XYalign [201] (Figure 4). Some others are based on the depth of coverage of the X and Y chromosome reads at selected genomic positions, such as that conducted by Somalier [200].
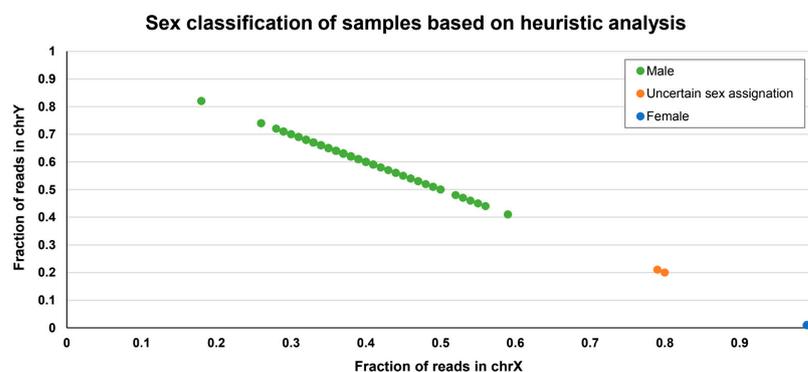
Figure 4. Typical sex inference of multiple samples based on read depth of X and Y chromosomes heuristic analysis. Uncertain sex assignation can aid the detection of sequencing errors in a multi-sample project [202].

6. Long-Read Sequencing Technologies in Cancer Genomics

A broad range of long-read sequencing technologies, also known as TGS, is flourishing and being used to improve our knowledge of complex regions and SVs that were difficult to resolve or were missed by short-read sequencing analysis [203]. Long-read sequencing is quickly evolving and becoming prevalent in cancer studies, allowing us to fully characterize novel somatic mutations involved in cancer, such as CNVs and SVs [204], and helping to identify further driver genes.

Several long-read sequencing technologies have been developed recently (Table 3). However, the prevailing TGS technologies are now the ones developed and marketed by ONT and PacBio. Other approaches, such as linked-reads [205], chromosome conformation capture sequencing (Hi-C) [206], and optical-mapping [207], are worth highlighting as well since they have demonstrated utility in cancer research.

Table 3. Main long-read sequencing technologies or approaches and highlighted cancer genomics research.

Technology	Instruments	Read Characteristics	Related Somatic Studies
Oxford Nanopore Technologies	MinION GridION PromethION	Single molecule reads, average read length ~15–20 Kb (max ~2 Mb), with an error rate of 5–10%	Brain tumor [208], lung cancer [209]
Pacific Biosciences	Sequel Sequel II	HiFi reads, average read length ~15–20 Kb (max ~65 Kb), with error rate of 1%	Breast cancer [19]
Linked-reads (10× Genomics)	NextSeq HiSeq NovaSeq	Linked-reads obtained from short reads, average length ~100 Kb	Prostate cancer [210], gastric cancer [211]
Hi-C	NextSeq HiSeq NovaSeq	~1 kb–1 Mb resolution, without base pair resolution	Pancreatic cancer [212]
Optical maps (BioNano Genomics)	NextSeq HiSeq NovaSeq	Optical mapping of long fragments, average length 250 Kb, without base pair resolution	Leukemia [213]

Hi-C, chromosome conformation capture sequencing; Kb, kilobases; Mb, megabases.

ONT has marketed a number of platforms such as the MinION, GridION, or the PromethION, among others, which are capable of sequencing ultra-long fragments of DNA or directly of RNA, offering also real-time analysis [214]. The MinION, released to the market in 2014 and the first device using nanopore technology, is a portable sequencer capable of sequencing whole small genomes or exomes, metagenomes, and transcriptomes [215,216]. GridION runs up to five parallel MinION flow cells and is suitable for medium-scale projects such as larger genomes, whole transcriptomes, or large numbers of samples. PromethION has 24 or 48 parallel flow cells and is a high-throughput device for large-scale projects, suitable for larger genomes and population sequencing. Lower-scale variations of PromethION are about to be released. These devices use flow cells that contain an array of multiple and parallel nanopores embedded into an electrically resistant polymer membrane. Single-stranded DNA or RNA molecules can pass through these nanopores with the help of proteins by means of an ionic current produced by applying a constant voltage. As nucleotides pass through, the current is disrupted, producing a characteristic change [217]. This signal is measured and then decoded using basecalling algorithms to determine the corresponding nucleotide type in real-time [218]. Theoretically, the length of the reads has no limits. However, in practice, the longest reads are currently capped at a maximum size of ~2 Mb imposed by the challenges of handling very long DNA molecules.

PacBio uses proprietary SMRT (single-molecule real-time) technology [219], a method based on a single-DNA polymerase attached in zero-mode waveguides (ZMWs), subwavelength optical nanostructures, to detect fluorescent signals. PacBio released the Sequel system in 2015, a platform that uses this sequencing process to obtain long-read sequences longer than 10 kb. In early 2019, PacBio released the Sequel II sequencing platform, an improvement of the Sequel I platform with higher data output. Moreover, together with Sequel II, they developed circular consensus sequencing (CCS), an improved method to

produce HiFi (high-fidelity) sequences with high base accuracy (>99%) in reads of about 10–15 kb in length. This makes PacBio technology suitable for de novo assembly, RNA sequencing, or comprehensive variant detection [220].

Another approach, known as linked-read technology [221], was developed by 10X Genomics and had a significant impact on the analysis in the determination of phased haplotypes and the identification of large genomic rearrangements. However, this sequencing method has been discontinued. New alternatives, such as the Hi-C technique [222], a method developed to study 3D genome folding, generate synthetic long-reads using short-reads while adding information from long DNA strands [223]. This methodology helps to build chromosome and genome structures, improving phasing and scaffolding to provide high-quality draft assemblies [224], helping to explain events such as genome folding or gene regulation [225]. It is also worth noting the utility of optical mapping technologies such as BioNano, which labels DNA sequences and then generates genome maps, to improve the assembly by scaffolding the assembled contigs [226] or to discover large SVs [227,228].

In 2020, the T2T announced the first gapless de novo assembly of a human X chromosome using ultra-long DNA reads from ONT sequencing of the T2T-CHM13 genome [229]. In 2022, combining several of these technologies, the T2T reported the first complete human genome, unraveling the last 8% of the genome that remained unresolved [111]. Despite there is yet a lack of melanoma studies leveraging these technologies [95], these will bring improvements to the field, by allowing the implementation of new reference databases, reducing variant calling errors, and improving genetic analyses of de novo and somatic mutations [112].

6.1. Advantages and Limitations of Long-Read Sequencing in Cancer Genomics

Long-read sequencing offers several advantages over short-read sequencing approaches. They are particularly useful in de novo assembly strategies because, by solving the low complexity and repetitive regions, they can reconstruct accurate and high-resolution genome assemblies [230]. This also has a positive effect on the ability to resolve repetitive and difficult to map regions [160], the comprehensive discovery of large and complex SVs, the full characterization of transcriptomes and the alternative splicing transcript species, the capacity for variant phasing in chromosomes, and the direct detection of epigenetic changes [231]. In the case of ONT, the MinION sequencing device allows to sequence samples in real-time and in a portable way [215]. All these benefits make this technology very promising for the characterization of any cancer type, as well as for developing specific therapeutic strategies [208].

As previously mentioned, TGS has several technical advantages and improvements compared to the traditional NGS. However, there still exist some limitations and challenges that are associated with these technologies, which may explain why their use in cancer research is not that widespread yet, particularly in the case of melanoma [95]. The main disadvantage of this sequencing technology is its higher error rate, which translates into a lower accuracy in the subsequent analysis. Nevertheless, base calling and error correction algorithms are continuously improving and progressively allowing to obtain improved read accuracies, which have a strong impact on read alignment and variant detection [232]. In addition, sample requirements for sequencing long-reads are substantially higher compared to short-read technologies. The amount of input DNA required nowadays is particularly problematic in tumor samples, as the starting material of these samples is usually limited and is typically degraded if the source is FFPE samples. In order to maximize the yield from sequencing and improve the quality of the data obtained, specific protocols for DNA extraction are necessary [233].

## 6.2. An exemplar Application of WGS with Long Reads from ONT

As discussed before, long-read sequencing platforms offer advantages for the de novo assembly of genomes compared to the classical NGS methods [234]. ONT and PacBio

have accelerated computing times and increased the read length up to several thousand base pairs, despite the higher error rate, making it possible to assess human genome variation from de novo assemblies [235]. Longer sequences make it easy to find overlaps with other sequences of the experiment, obtaining better results by correctly assembling DNA fragments and facilitating the spanning over repetitive genomics regions (as gene duplications, transposons, or satellites). Because of these improvements, it is possible to run de novo assembly to obtain a personalized genome and use it as a reference for detecting somatic events from cancers [236]. Using a personalized genome assembly [237], rather than the standard reference such as GRCh37 or GRCh38, as a reference for tumor samples could improve read alignment and somatic mutation discovery.

6.2.1. Library Preparation and Sequencing

ONT provides a comprehensive range of DNA and RNA library preparation kits, offering high throughput with low DNA input, fast modes for library preparation, and long reads. This technology has a wide variety of solutions, including WGS, targeted sequencing, and RNA sequencing, among many others. Whereas with other sequencing technologies the read length is limited by the technology itself, in the case of long reads this limitation is given by the quality and size of the starting DNA material. Thus, the challenge is in how to improve DNA extraction to preserve DNA purity and integrity. For analyzing human cancer WGS, one can proceed with the ONT ligation-based library preparation kit (SQK-LSK110) and a PromethION platform, which can generate up to 150 Gb of sequence per flow cell and offers $30\times$ coverage WGS for less than 1000 dollars [238].

Signals from the sequencing process are stored within FAST5 format files and could be processed by basecalling algorithms, such as Guppy or Bonito (ONT), to decode the sequence into FASTQ files. Basecalled reads can be inspected using tools such as pycoQC [239] (Figure 5) or NanoPlot [240] to generate interactive QC metrics and plots.
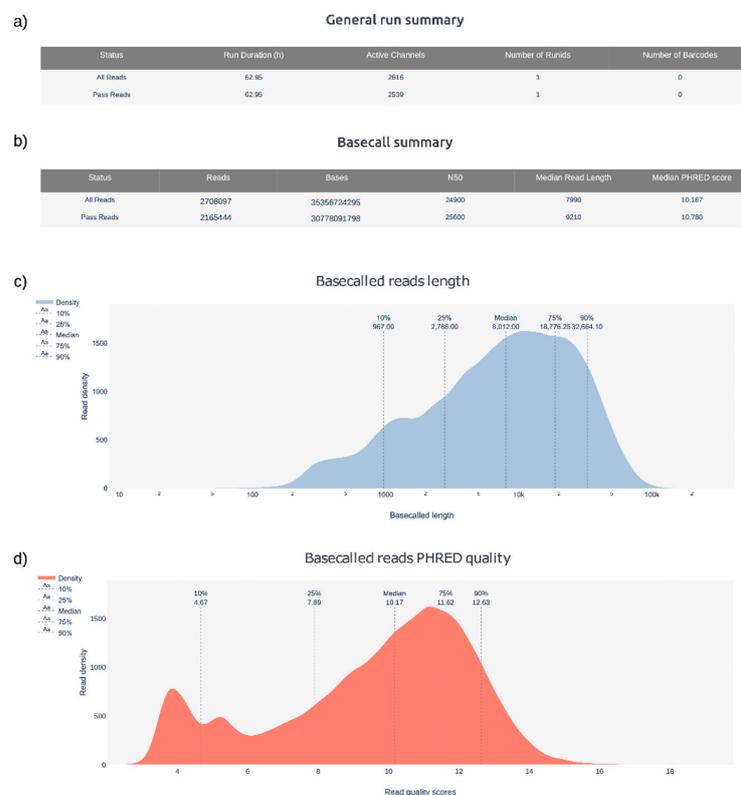
Figure 5. Summary report of ONT basecalling generated by means of pycoQC [239]. (**a**) General run statistics; (**b**) basecalling summary: number of total reads and bases, N50, median read length and median score; (**c**) plot of basecalled reads length; (**d**) plot of basecalled reads Phred scores.

6.2.2. Bioinformatic Tools for Long-Read Analysis

In recent years, the research activity involving long-read technologies has grown rapidly [238,241]. This effect produces an exponential development of bioinformatic tools for long-read sequence analysis. Some of those tools are maintained by companies in their own repositories, while others are open-source applications and pipelines developed by researchers or laboratories. Table 4 provides some of the most common computational tools and applications for long-read sequencing data analysis, including basecalling, error correction, and de novo assembly, among others.

Table 4. Common computational tools for long-read sequencing data analysis.

Bioinformatic Analysis	Tool	Sequencing Strategy	References
Base calling	Guppy, Bonito	ONT	https://github.com/nanoporetech/ (accessed on 2 August 2022)
	Generate CCS	PacBio	[242]
Quality control	pycoQC, NanoPack	ONT	[239,240]
	Isoseq3	PacBio	https://github.com/PacificBiosciences/IsoSeq (accessed on 2 August 2022)
Read-error correction	Canu	ONT	[243]
	LoRMA	PacBio	[244]
DNA methylation	pycoMeth, DeepSignal, Megalodon	ONT	[245,246]; https://github.com/nanoporetech/megalodon (accessed on 2 August 2022)
	pb-CpG-tools	PacBio	https://github.com/PacificBiosciences/pb-CpG-tools (accessed on 2 August 2022)
Alignment	minimap2, NGMLR	ONT	[117,247]
	pbmm2	PacBio	https://github.com/PacificBiosciences/pbmm2 (accessed on 2 August 2022)
SNV calling	Longshot, DeepVariant	ONT, PacBio	[248,249]
SV calling	Sniffles, SVIM, SVIM-asm, cuteSV	ONT	[247,250–252]
	pbsv	PacBio	https://github.com/PacificBiosciences/pbsv (accessed on 2 August 2022)
De novo assembly	Flye, Shasta	ONT	[253,254]
	Hifiasm, FALCON	PacBio	[255,256]
Hybrid assembly	MaSuRCA, WENGAN	ONT, PacBio	[257,258]
Polishing	Racon, Medaka, Pilon	ONT	[259,260]
	Pilon, Quiver, Arrow	PacBio	[260,261]

SNV, small nucleotide variant; SV, structural variant; ONT, Oxford Nanopore Technologies; PacBio, Pacific Biosciences.

6.2.3. De Novo Genome Assembly

De novo assembly of genomes is one of the main uses of the long reads generated by ONT and PacBio sequencing. The main goal of de novo assembly is to attempt to reconstruct the whole genome sequence without information from a reference genome [224,262] or in situations that aim to avoid potential biases from using the standard reference genome [237,263,264] that could be incomplete or not be representative of the population of the study [226,265,266]. De novo assembly algorithms manage to build contiguous and accurate sequences that represent the genome of the analyzed individual without information on the structure and complexity of the genome. There are two main algorithmic

strategies to run de novo assembly, one based on building a de Bruijn graph (DBG) and the other based on the overlap layout consensus (OLC) algorithm [267,268]. A basic workflow to complete a de novo genome assembly using Flye [253] as the assembler and using both long and short reads to run several polishing steps is shown in Figure 6.
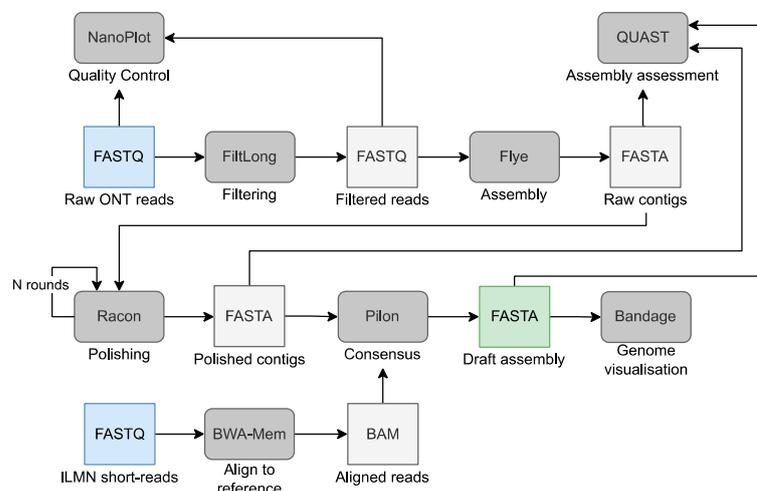
Figure 6. Basic de novo genome assembly pipeline using long reads from ONT. This pipeline starts by filtering raw ONT long reads using FiltLong to discard short and low-quality reads. NanoPlot is then used to check the quality of both raw and filtered reads. The remaining reads are de novo assembled by using Flye. Resulting contigs from assembly are polished with multiple rounds of Racon and Pilon, in case that short reads of the same sample are available. Resulting assemblies can be assessed using QUAST and visualized with Bandage.

The quality of the assemblies could be improved by combining data from different sequencing technologies, such as long reads from ONT and short reads from Illumina, Inc. With this approach, tools such as MaSuRCA [257] run hybrid de novo assemblies, using the long reads to scaffold contigs generated by short reads to solve regions that cannot be resolved using short reads alone [269].

6.2.4. SV Calling

Identification of small variants (SNVs and indels <50 bp) is essentially well resolved with short reads. However, the difficulty increases when it comes to detecting larger sequence mutations with this technology [157], and it is especially complex when dealing with somatic variants due to the purity and heterogeneity of tumor samples. These limitations have hindered the study of SVs in the past, despite their interest in cancer. Nowadays, powered by TGS technologies and the ability of long reads to cover repetitive elements and large and complex regions, the detection of SVs is more manageable by calling tools [270] and allows a more precise characterization of the SVs of the human genome [271]. Nevertheless, two of the previously mentioned disadvantages must be considered for alignment and SV calling the following steps: the high error rate and the throughput (translated into lower coverage) compared to short-read sequencing.

SV methods with long reads are mainly focused on the following two strategies: one based on aligning reads to a reference and another based on de novo assembly. Approaches that use de novo assembly are based on assembling reads to longer sequences, namely, contigs or scaffolds, and discovering SVs by comparing aligned reads to the assembled sequences and to a reference [251]. On the contrary, read alignment approaches are based on aligning raw reads against a reference and analyzing the resulting alignments to detect SVs.

A large number of tools rely on these methods, and their comparison helps in the selection of the optimal tool and the strategy for SV calling [272]. The application of long-

read sequencing technologies to study SVs has increased recently in cancer studies [19,273], although there are still very limited examples of this in melanoma research [274].

7. Discussion

NGS platforms have made a huge impact in the characterization of human tumors, helping to understand and identify several types of cancer and establish new targeted treatments. Moreover, despite the technical challenges that are due to the quality and quantity of the tumor samples, advances in sample preparation methods have enabled the full characterization of cancer genomes, transcriptomes, and epigenomes [275]. These advances are allowing the improvement our understanding of cancer-specific small-sequence mutations, such as SNVs and indels, and large genetic variations, such as CNVs or structural rearrangements.

Bioinformatics pipelines for the analysis and clinical interpretation of cancer genomic results have been implemented in many platforms and laboratories. Likewise, multiple computational tools have been developed for the analysis of oncological NGS data. Most of these tools are used through the command line, sometimes being complex to parameterize and optimize. The choice and application of these tools depend on the characteristics of each specific project. Thus, the optimal configuration must be determined empirically for the sequencing strategy, the sample type, and the computational resources.

Advances in NGS technologies, improvements in the variant detection algorithms, and further development of specific human cancer databases and functional annotations of the genetic variants, as well as the reduction of the cost of sequencing, will impact the field of cancer genomics, bolstering the development of better treatments [276]. All of these, combined with the efforts from international research groups and consortia, such as the TCGA or the ICGC, will provide new and better insights into the genetic characteristics of diverse types of cancer. Integrated data analysis is an important aspect of precision oncology research and has led to groundbreaking discoveries that would not have been possible without multi-omics analysis. Although various efforts have been made to develop machine learning-based methods to automate the integration levels of omics data [277], this is not a simple task. The approaches still need to solve problems such as batch effects and normalization within the integration analysis of the different types of data, and also be able to integrate different types of data, such as metabolomics data, which have shown to have a significant impact on cancer pathogenesis [278].

Optimal annotation and prioritization of variants is also a bottleneck in current precision oncology. This process requires databases containing curated variants as well as links with mechanistic effects and potential drug interaction data. However, simple and completely automated methods to carry out these processes are not yet available. In addition, large datasets of cancer patients, including their response to therapy, are needed so that effective machine-learning-based algorithms can be designed.

Long-read sequencing enables more comprehensive analysis of cancer genomes, solves complex genomic aberrations, improves the study of long transcript isoforms, and epigenomic modifications. The use of long reads in a multitude of cancer types is widespread, despite it not yet being leveraged in cutaneous melanoma. However, this technology still has several limitations to solve, including the higher error rates and the difficulty of obtaining high-molecular weight DNA material in sufficient quantities from the most common types of biobanked tumor samples [279].

8. Concluding Remarks

Because of the multiple advantages, the tremendous growth of applications, and the steep reductions in the cost per base, NGS has revolutionized research in melanoma and cancer genomics in the past decade. Despite the recent advances in the technology, including moving towards leveraging longer reads to optimally assess SVs, which are central in cancer, these improvements still lack a translation to the field. In order to continue pushing this field further, it will also be important to standardize the bioinformatics procedures

and update the annotation databases with the most recent and complete references, as they will allow us to better identify somatic and germline sequence variation that is key in the pathogenesis of melanoma.

References

1. Ottaviano, M.; Giunta, E.; Tortora, M.; Curvietto, M.; Attademo, L.; Bosso, D.; Cardalesi, C.; Rosanova, M.; De Placido, P.; Pietroluongo, E.; et al. BRAF Gene and Melanoma: Back to the Future. *Int. J. Mol. Sci.* **2021**, *22*, 3474. [CrossRef] [PubMed]
2. Walia, V.; Mu, E.W.; Lin, J.C.; Samuels, Y. Delving into somatic variation in sporadic melanoma. *Pigment Cell Melanoma Res.* **2012**, *25*, 155–170. [CrossRef] [PubMed]
3. Garbe, C.; Peris, K.; Hauschild, A.; Saiag, P.; Middleton, M.; Bastholt, L.; Grob, J.-J.; Malvehy, J.; Newton-Bishop, J.; Stratigos, A.J.; et al. Diagnosis and treatment of melanoma. European consensus-based interdisciplinary guideline—Update 2016. *Eur. J. Cancer* **2016**, *63*, 201–217. [CrossRef] [PubMed]
4. Gandini, S.; Sera, F.; Cattaruzza, M.S.; Pasquini, P.; Abeni, D.; Boyle, P.; Melchi, C.F. Meta-analysis of risk factors for cutaneous melanoma: I. Common and atypical naevi. *Eur. J. Cancer* **2005**, *41*, 28–44. [CrossRef]
5. Bataille, V.; Bishop, J.A.; Sasieni, P.; Swerdlow, A.; Pinney, E.; Griffiths, K.M.; Cuzick, J. Risk of cutaneous melanoma in relation to the numbers, types and sites of naevi: A case-control study. *Br. J. Cancer* **1996**, *73*, 1605–1611. [CrossRef] [PubMed]
6. Krengel, S.; Hauschild, A.; Schäfer, T. Melanoma risk in congenital melanocytic naevi: A systematic review. *Br. J. Dermatol.* **2006**, *155*, 1–8. [CrossRef]
7. Gandini, S.; Sera, F.; Cattaruzza, M.S.; Pasquini, P.; Picconi, O.; Boyle, P.; Melchi, C.F. Meta-analysis of risk factors for cutaneous melanoma: II. Sun exposure. *Eur. J. Cancer* **2005**, *41*, 45–60. [CrossRef]
8. Greinert, R. Skin Cancer: New Markers for Better Prevention. *Pathobiology* **2009**, *76*, 64–81. [CrossRef]
9. Jemal, A.; Devesa, S.S.; Fears, T.R.; Hartge, P. Cancer surveillance series: Changing patterns of cutaneous malignant melanoma mortality rates among whites in the United States. *JNCI J. Natl. Cancer Inst.* **2000**, *92*, 811–818. [CrossRef]
10. Tucker, M.A. Is Sunlight Important to Melanoma Causation? *Cancer Epidemiol. Biomark. Prev.* **2008**, *17*, 467–468. [CrossRef]
11. Mitra, D.; Luo, X.; Morgan, A.; Wang, J.; Hoang, M.P.; Lo, J.; Guerrero, C.R.; Lennerz, J.K.; Mihm, M.C.; Wargo, J.A.; et al. An ultraviolet-radiation-independent pathway to melanoma carcinogenesis in the red hair/fair skin background. *Nature* **2012**, *491*, 449–453. [CrossRef] [PubMed]
12. Raimondi, S.; Sera, F.; Gandini, S.; Iodice, S.; Caini, S.; Maisonneuve, P.; Fargnoli, M.C. MC1R variants, melanoma and red hair color phenotype: A meta-analysis. *Int. J. Cancer* **2008**, *122*, 2753–2760. [CrossRef]
13. Flaherty, K.T.; Puzanov, I.; Kim, K.B.; Ribas, A.; McArthur, G.A.; Sosman, J.A.; O'Dwyer, P.J.; Lee, R.J.; Grippo, J.F.; Nolop, K.; et al. Inhibition of Mutated, Activated BRAF in Metastatic Melanoma. *N. Engl. J. Med.* **2010**, *363*, 809–819. [CrossRef] [PubMed]
14. Jakob, J.A.; Bassett, R.L.; Ng, C.S.; Curry, J.L.; Joseph, R.; Alvarado, G.C.; Apn, M.L.R.; Richard, J.; Gershenwald, J.E.; Kim, K.B.; et al. NRAS mutation status is an independent prognostic factor in metastatic melanoma. *Cancer* **2011**, *118*, 4014–4023. [CrossRef] [PubMed]

15. Curtin, J.A.; Busam, K.; Pinkel, D.; Bastian, B.C. Somatic Activation of KIT in Distinct Subtypes of Melanoma. *J. Clin. Oncol.* **2006**, *24*, 4340–4346. [CrossRef]

16. Krauthammer, M.; Kong, Y.; Bacchiocchi, A.; Evans, P.; Pornputtapong, N.; Wu, C.; McCusker, J.P.; Ma, S.; Cheng, E.; Straub, R.; et al. Exome sequencing identifies recurrent mutations in NF1 and RASopathy genes in sun-exposed melanomas. *Nat. Genet.* **2015**, *47*, 996–1002. [CrossRef]

17. Rossi, M.; Pellegrini, C.; Cardelli, L.; Ciciarelli, V.; di Nardo, L.; Fargnoli, M.C. Familial melanoma: Diagnostic and management implications. *Dermatol. Pract. Concept.* **2019**, *9*, 10–16. [CrossRef]

18. Rusch, M.; Nakitandwe, J.; Shurtleff, S.; Newman, S.; Zhang, Z.; Edmonson, M.N.; Parker, M.; Jiao, Y.; Ma, X.; Liu, Y.; et al. Clinical cancer genomic profiling by three-platform sequencing of whole genome, whole exome and transcriptome. *Nat. Commun.* **2018**, *9*, 1–13. [CrossRef]

19. Nattestad, M.; Goodwin, S.; Ng, K.; Baslan, T.; Sedlazeck, F.J.; Rescheneder, P.; Garvin, T.; Fang, H.; Gurtowski, J.; Hutton, E.; et al. Complex rearrangements and oncogene amplifications revealed by long-read DNA and RNA sequencing of a breast cancer cell line. *Genome Res.* **2018**, *28*, 1126–1135. [CrossRef]

20. Suzuki, A.; Suzuki, M.; Mizushima-Sugano, J.; Frith, M.; Makałowski, W.; Kohno, T.; Sugano, S.; Tsuchihara, K.; Suzuki, Y. Sequencing and phasing cancer mutations in lung cancers using a long-read portable sequencer. *DNA Res.* **2017**, *24*, 585–596. [CrossRef]

21. Vanni, I.; Tanda, E.T.; Spagnolo, F.; Andreotti, V.; Bruno, W.; Ghiorzo, P. The Current State of Molecular Testing in the BRAF-Mutated Melanoma Landscape. *Front. Mol. Biosci.* **2020**, *7*, 113. [CrossRef] [PubMed]

22. Lightbody, G.; Haberland, V.; Browne, F.; Taggart, L.; Zheng, H.; Parkes, E.; Blayney, J.K. Review of applications of high-throughput sequencing in personalized medicine: Barriers and facilitators of future progress in research and clinical application. *Brief. Bioinform.* **2019**, *20*, 1795–1811. [CrossRef] [PubMed]

23. Tokuda, Y.; Nakamura, T.; Satonaka, K.; Maeda, S.; Doi, K.; Baba, S.; Sugiyama, T. Fundamental study on the mechanism of DNA degradation in tissues fixed in formaldehyde. *J. Clin. Pathol.* **1990**, *43*, 748–751. [CrossRef] [PubMed]

24. Do, H.; Dobrovic, A. Sequence Artifacts in DNA from Formalin-Fixed Tissues: Causes and Strategies for Minimization. *Clin. Chem.* **2015**, *61*, 64–71. [CrossRef]

25. Eckhart, L.; Bach, J.; Ban, J.; Tschachler, E. Melanin Binds Reversibly to Thermostable DNA Polymerase and Inhibits Its Activity. *Biochem. Biophys. Res. Commun.* **2000**, *271*, 726–730. [CrossRef] [PubMed]

26. Guyard, A.; Boyez, A.; Pujals, A.; Robe, C.; Van Nhieu, J.T.; Allory, Y.; Moroch, J.; Georges, O.; Fournet, J.-C.; Zafrani, E.-S.; et al. DNA degrades during storage in formalin-fixed and paraffin-embedded tissue blocks. *Virchows Arch.* **2017**, *471*, 491–500. [CrossRef]

27. Ludyga, N.; Grünwald, B.; Azimzadeh, O.; Englert, S.; Höfler, H.; Tapio, S.; Aubele, M. Nucleic acids from long-term preserved FFPE tissues are suitable for downstream analyses. *Virchows Arch.* **2012**, *460*, 131–140. [CrossRef]

28. Millán-Esteban, D.; Reyes-García, D.; García-Casado, Z.; Bañuls, J.; López-Guerrero, J.A.; Requena, C.; Rodríguez-Hernández, A.; Traves, V.; Nagore, E. Suitability of melanoma FFPE samples for NGS libraries: Time and quality thresholds for downstream molecular tests. *BioTechniques* **2018**, *65*, 79–85. [CrossRef]

29. Mathieson, W.; Thomas, G.A. Why Formalin-fixed, Paraffin-embedded Biospecimens Must Be Used in Genomic Medicine: An Evidence-based Review and Conclusion. *J. Histochem. Cytochem.* **2020**, *68*, 543–552. [CrossRef]

30. van Dijk, E.L.; Jaszczyszyn, Y.; Thermes, C. Library preparation methods for next-generation sequencing: Tone down the bias. *Exp. Cell Res.* **2014**, *322*, 12–20. [CrossRef]

31. Petty, D.R.; Hassan, O.A.; Barker, C.S.; O'Neill, S.S. Rapid BRAF Mutation Testing in Pigmented Melanomas. *Am. J. Dermatopathol.* **2019**, *42*, 343–348. [CrossRef] [PubMed]

32. Frouin, E.; Maudelonde, T.; Senal, R.; Larrieux, M.; Costes, V.; Godreuil, S.; Vendrell, J.A.; Solassol, J. Comparative Methods to Improve the Detection of BRAF V600 Mutations in Highly Pigmented Melanoma Specimens. *PLoS ONE* **2016**, *11*, e0158698. [CrossRef] [PubMed]

33. Vicente, A.L.S.A.; Bianchini, R.A.; Laus, A.C.; Macedo, G.; Reis, R.M.; de Lima Vazquez, V. Comparison of protocols for removal of melanin from genomic DNA to optimize PCR amplification of DNA purified from highly pigmented lesions. *Histol. Histopathol. Cell Biol. Tissue Eng.* **2019**, *34*, 1089–1096. [CrossRef]

34. Kukurba, K.R.; Montgomery, S.B. RNA Sequencing and Analysis. *Cold Spring Harb. Protoc.* **2015**, *2015*, 951–969. [CrossRef]

35. Lu, W.; Zhou, Q.; Chen, Y. Impact of RNA degradation on next-generation sequencing transcriptome data. *Genomics* **2022**, *114*, 110429. [CrossRef]

36. Rudloff, U.; Bhanot, U.; Gerald, W.; Klimstra, D.S.; Jarnagin, W.R.; Brennan, M.; Allen, P.J. Biobanking of Human Pancreas Cancer Tissue: Impact of Ex-Vivo Procurement Times on RNA Quality. *Ann. Surg. Oncol.* **2010**, *17*, 2229–2236. [CrossRef]

37. Völler, D.; Reinders, J.; Meister, G.; Bosserhoff, A.-K. Strong reduction of AGO2 expression in melanoma and cellular consequences. *Br. J. Cancer* **2013**, *109*, 3116–3124. [CrossRef]

38. Sanger, F.; Nicklen, S.; Coulson, A.R. DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA* **1977**, *74*, 5463–5467. [CrossRef]

39. Maria, M.; Ajmal, M.; Azam, M.; Waheed, N.K.; Siddiqui, S.N.; Mustafa, B.; Ayub, H.; Ali, L.; Ahmad, S.; Micheal, S.; et al. Homozygosity Mapping and Targeted Sanger Sequencing Reveal Genetic Defects Underlying Inherited Retinal Disease in Families from Pakistan. *PLoS ONE* **2015**, *10*, e0119806. [CrossRef]

40.	Bezdíčka, M.; Štolbová; Seeman, T.; Cinek, O.; Malina, M.; Šimánková, N.; Průhová, .; Zieg, J. Genetic diagnosis of steroid-resistant nephrotic syndrome in a longitudinal collection of Czech and Slovak patients: A high proportion of causative variants in NUP93. *Pediatr. Nephrol.* **2018**, *33*, 1347–1363. [CrossRef]

41.	Liang, C.; Wu, Z.; Gan, X.; Liu, Y.; You, Y.; Liu, C.; Zhou, C.; Liang, Y.; Mo, H.; Chen, A.M.; et al. Detection of Rare Mutations in *EGFR*-ARMS-PCR-Negative Lung Adenocarcinoma by Sanger Sequencing. *Yonsei Med. J.* **2018**, *59*, 13–19. [CrossRef] [PubMed]

42.	Bruijns, B.; Tiggelaar, R.M.; Gardeniers, J. Massively parallel sequencing techniques for forensics: A review. *Electrophoresis* **2018**, *39*, 2642–2654. [CrossRef] [PubMed]

43.	International Human Genome Sequencing Consortium. International Human Genome Sequencing Consortium Finishing the euchromatic sequence of the human genome. *Nature* **2004**, *431*, 931–945. [CrossRef]

44.	Sheen, Y.-S.; Liao, Y.-H.; Liau, J.-Y.; Lin, M.-H.; Hsieh, Y.-C.; Jee, S.-H.; Chu, C.-Y. Prevalence of BRAF and NRAS mutations in cutaneous melanoma patients in Taiwan. *J. Formos. Med. Assoc.* **2016**, *115*, 121–127. [CrossRef]

45.	Ren, M.; Zhang, J.; Kong, Y.; Bai, Q.; Qi, P.; Wang, Q.; Zhou, X.; Chen, Y.; Zhu, X. BRAF, C-KIT, and NRAS mutations correlated with different clinicopathological features: An analysis of 691 melanoma patients from a single center. *Ann. Transl. Med.* **2022**, *10*, 31. [CrossRef]

46.	Fatnassi-Mersni, G.; Arfaoui, A.T.; Cherni, M.; Jones, M.; Zeglaoui, F.; Ouzari, H.I.; Rammeh, S. Molecular and immunohistochemical analysis of BRAF gene in primary cutaneous melanoma: Discovery of novel mutations. *J. Cutan. Pathol.* **2020**, *47*, 794–799. [CrossRef]

47.	Cheng, L.Y.; Haydu, L.E.; Song, P.; Nie, J.; Tetzlaff, M.T.; Kwong, L.N.; Gershenwald, J.E.; Davies, M.A.; Zhang, D.Y. High sensitivity sanger sequencing detection of BRAF mutations in metastatic melanoma FFPE tissue specimens. *Sci. Rep.* **2021**, *11*, 1–9. [CrossRef]

48.	Chen, C.; Fangxuqian, X.; Sun, S. Diagnosis of polyglutamine spinocerebellar ataxias by polymerase chain reaction amplification and Sanger sequencing. *Mol. Med. Rep.* **2018**, *18*, 1037–1042. [CrossRef]

49.	Fukuta, M.; Gaballah, M.; Takada, K.; Miyazaki, H.; Kato, H.; Aoki, Y.; Hamed, S.S.; ElMorsi, D.A.; ElDakroory, S.A. Genetic polymorphism of 27 X-chromosomal short tandem repeats in an Egyptian population. *Leg. Med.* **2019**, *37*, 64–66. [CrossRef]

50.	Khan, A.A.; Perveen, R.; Sheikh, N.; Abbasi, B.H.A.; Batool, Z.; Shahzad, M.; Kaleem, S. Genetic polymorphism of 15 autosomal short tandem repeats in Baloch population of Pakistan. *Int. J. Legal Med.* **2018**, *133*, 775–776. [CrossRef]

51.	Nyren, P.; Pettersson, B.; Uhlen, M. Solid Phase DNA Minisequencing by an Enzymatic Luminometric Inorganic Pyrophosphate Detection Assay. *Anal. Biochem.* **1993**, *208*, 171–175. [CrossRef]

52.	Edlundhrose, E.; Egyhazi, S.; Omholt, K.; Manssonbrahme, E.; Platz, A.; Hansson, J.; Lundeberg, J. NRAS and BRAF mutations in melanoma tumours in relation to clinical characteristics: A study based on mutation screening by pyrosequencing. *Melanoma Res.* **2006**, *16*, 471–478. [CrossRef] [PubMed]

53.	Yaman, B.; Kandiloğlu, G.; Akalin, T. BRAF-V600 Mutation Heterogeneity in Primary and Metastatic Melanoma. *Am. J. Dermatopathol.* **2016**, *38*, 113–120. [CrossRef]

54.	Margulies, M.; Egholm, M.; Altman, W.E.; Attiya, S.; Bader, J.S.; Bemben, L.A.; Berka, J.; Braverman, M.S.; Chen, Y.-J.; Chen, Z.; et al. Genome sequencing in microfabricated high-density picolitre reactors. *Nature* **2005**, *437*, 376–380. [CrossRef] [PubMed]

55.	Srivastava, S.; Cohen, J.S.; Vernon, H.; Barañano, K.; McClellan, R.; Jamal, L.; Naidu, S.; Fatemi, A. Clinical whole exome sequencing in child neurology practice. *Ann. Neurol.* **2014**, *76*, 473–483. [CrossRef] [PubMed]

56.	Yang, Y.; Muzny, D.M.; Xia, F.; Niu, Z.; Person, R.; Ding, Y.; Ward, P.; Braxton, A.; Wang, M.; Buhay, C.; et al. Molecular Findings Among Patients Referred for Clinical Whole-Exome Sequencing. *JAMA* **2014**, *312*, 1870–1879. [CrossRef] [PubMed]

57.	Vissers, L.E.; van Nimwegen, K.J.; Schieving, J.H.; Kamsteeg, E.-J.; Kleefstra, T.; Yntema, H.G.; Pfundt, R.; van der Wilt, G.J.; Krabbenborg, L.; Brunner, H.G.; et al. A clinical utility study of exome sequencing versus conventional genetic testing in pediatric neurology. *Genet. Med.* **2017**, *19*, 1055–1063. [CrossRef]

58.	Shendure, J.; Porreca, G.J.; Reppas, N.B.; Lin, X.; McCutcheon, J.P.; Rosenbaum, A.M.; Wang, M.D.; Zhang, K.; Mitra, R.D.; Church, G.M. Accurate Multiplex Polony Sequencing of an Evolved Bacterial Genome. *Science* **2005**, *309*, 1728–1732. [CrossRef]

59.	Braslavsky, I.; Hebert, B.; Kartalov, E.; Quake, S.R. Sequence information can be obtained from single DNA molecules. *Proc. Natl. Acad. Sci. USA* **2003**, *100*, 3960–3964. [CrossRef]

60.	Hintzsche, J.D.; Gorden, N.T.; Amato, C.M.; Kim, J.; Wuensch, K.E.; Robinson, S.E.; Applegate, A.J.; Couts, K.L.; Medina, T.M.; Wells, K.R.; et al. Whole-exome sequencing identifies recurrent SF3B1 R625 mutation and comutation of NF1 and KIT in mucosal melanoma. *Melanoma Res.* **2017**, *27*, 189–199. [CrossRef]

61.	Vergara, I.A.; Mintoff, C.P.; Sandhu, S.; McIntosh, L.; Young, R.J.; Wong, S.Q.; Colebatch, A.; Cameron, D.L.; Kwon, J.L.; Wolfe, R.; et al. Evolution of late-stage metastatic melanoma is dominated by aneuploidy and whole genome doubling. *Nat. Commun.* **2021**, *12*, 1–15. [CrossRef]

62.	Abdel-Rahman, M.H.; Sample, K.M.; Pilarski, R.; Walsh, T.; Grosel, T.; Kinnamon, D.; Boru, G.; Massengill, J.B.; Schoenfield, L.; Kelly, B.; et al. Whole Exome Sequencing Identifies Candidate Genes Associated with Hereditary Predisposition to Uveal Melanoma. *Ophthalmology* **2019**, *127*, 668–678. [CrossRef] [PubMed]

63.	Cai, H.; Jing, C.; Chang, X.; Ding, D.; Han, T.; Yang, J.; Lu, Z.; Hu, X.; Liu, Z.; Wang, J.; et al. Mutational landscape of gastric cancer and clinical application of genomic profiling based on target next-generation sequencing. *J. Transl. Med.* **2019**, *17*, 189. [CrossRef] [PubMed]

64. Zhang, J.; Walsh, M.F.; Wu, G.; Edmonson, M.N.; Gruber, T.A.; Easton, J.; Hedges, D.; Ma, X.; Zhou, X.; Yergeau, D.A.; et al. Germline Mutations in Predisposition Genes in Pediatric Cancer. *N. Engl. J. Med.* **2015**, *373*, 2336–2346. [CrossRef] [PubMed]

65. Reiman, A.; Kikuchi, H.; Scocchia, D.; Smith, P.; Tsang, Y.W.; Snead, D.; Cree, I.A. Validation of an NGS mutation detection panel for melanoma. *BMC Cancer* **2017**, *17*, 150. [CrossRef]

66. Newell, F.; Wilmott, J.S.; Johansson, P.A.; Nones, K.; Addala, V.; Mukhopadhyay, P.; Broit, N.; Amato, C.M.; Van Gulick, R.; Kazakoff, S.H.; et al. Whole-genome sequencing of acral melanoma reveals genomic complexity and diversity. *Nat. Commun.* **2020**, *11*, 1–14. [CrossRef] [PubMed]

67. Usera, A.D.-D.; Lorenzo-Salazar, J.M.; Rubio-Rodríguez, L.A.; Muñoz-Barrera, A.; Guillen-Guio, B.; Marcelino-Rodríguez, I.; García-Olivares, V.; Mendoza-Alvarez, A.; Corrales, A.; Íñigo-Campos, A.; et al. Evaluation of Whole-Exome Enrichment Solutions: Lessons from the High-End of the Short-Read Sequencing Scale. *J. Clin. Med.* **2020**, *9*, 3656. [CrossRef] [PubMed]

68. Jain, A.; Govindaraj, G.M.; Edavazhippurath, A.; Faisal, N.; Bhoyar, R.C.; Gupta, V.; Uppuluri, R.; Manakkad, S.P.; Kashyap, A.; Kumar, A.; et al. Whole genome sequencing identifies novel structural variant in a large Indian family affected with X-linked agammaglobulinemia. *PLoS ONE* **2021**, *16*, e0254407. [CrossRef]

69. Hou, Y.C.; Neidich, J.A.; Duncavage, E.J.; Spencer, D.H.; Schroeder, M.C. Clinical whole-genome sequencing in cancer diagnosis. *Hum. Mutat.* **2022**, *43*, 1519–1530. [CrossRef]

70. Li, Y.; PCAWG Structural Variation Working Group; Roberts, N.D.; Wala, J.A.; Shapira, O.; Schumacher, S.E.; Kumar, K.; Khurana, E.; Waszak, S.; Korbel, J.O.; et al. Patterns of somatic structural variation in human cancer genomes. *Nature* **2020**, *578*, 112–121. [CrossRef]

71. Cosenza, M.R.; Rodriguez-Martin, B.; Korbel, J.O. Structural Variation in Cancer: Role, Prevalence, and Mechanisms. *Annu. Rev. Genom. Hum. Genet.* **2022**, *23*, 123–152. [CrossRef] [PubMed]

72. Takai, E.; Nakamura, H.; Chiku, S.; Kubo, E.; Ohmoto, A.; Totoki, Y.; Shibata, T.; Higuchi, R.; Yamamoto, M.; Furuse, J.; et al. Whole-exome Sequencing Reveals New Potential Susceptibility Genes for Japanese Familial Pancreatic Cancer. *Ann. Surg.* **2020**, *275*, e652–e658. [CrossRef] [PubMed]

73. Liu, J.; Mao, R.; Ren, G.; Liu, X.; Zhang, Y.; Wang, J.; Wang, Y.; Li, M.; Qiu, Q.; Wang, L.; et al. Whole Exome Sequencing Identifies Putative Predictors of Recurrent Prostate Cancer with High Accuracy. *OMICS A J. Integr. Biol.* **2019**, *23*, 380–388. [CrossRef]

74. Chen, J.; Li, Y.; Wu, J.; Liu, Y.; Kang, S. Whole-exome sequencing reveals potential germline and somatic mutations in 60 malignant ovarian germ cell tumors. *Biol. Reprod.* **2021**, *105*, 164–178. [CrossRef] [PubMed]

75. Skopelitou, D.; Miao, B.; Srivastava, A.; Kumar, A.; Kuświk, M.; Dymerska, D.; Paramasivam, N.; Schlesner, M.; Lubinski, J.; Hemminki, K.; et al. Whole Exome Sequencing Identifies *APCDD1* and *HDAC5* Genes as Potentially Cancer Predisposing in Familial Colorectal Cancer. *Int. J. Mol. Sci.* **2021**, *22*, 1837. [CrossRef] [PubMed]

76. Yang, Y.; Gu, X.; Li, Z.; Zheng, C.; Wang, Z.; Zhou, M.; Chen, Z.; Li, M.; Li, D.; Xiang, J. Whole-exome sequencing of rectal cancer identifies locally recurrent mutations in the Wnt pathway. *Aging* **2021**, *13*, 23262–23283. [CrossRef] [PubMed]

77. Hayward, N.K.; Wilmott, J.S.; Waddell, N.; Johansson, P.A.; Field, M.A.; Nones, K.; Patch, A.-M.; Kakavand, H.; Alexandrov, L.B.; Burke, H.; et al. Whole-genome landscapes of major melanoma subtypes. *Nature* **2017**, *545*, 175–180. [CrossRef]

78. Wilmott, J.S.; Johansson, P.A.; Newell, F.; Waddell, N.; Ferguson, P.; Quek, C.; Patch, A.-M.; Nones, K.; Shang, P.; Pritchard, A.L.; et al. Whole genome sequencing of melanomas in adolescent and young adults reveals distinct mutation landscapes and the potential role of germline variants in disease susceptibility. *Int. J. Cancer* **2018**, *144*, 1049–1060. [CrossRef]

79. Araújo, G.; Marinho, A.N.R.; Anaissi, A.K.; Vinasco-Sandoval, T.; Ribeiro-Dos-Santos, A.; Vidal, A.; De Araújo, G.S.; Demachki, S.; Ribeiro-Dos-Santos, Â. Whole mitochondrial genome sequencing highlights mitochondrial impact in gastric cancer. *Sci. Rep.* **2019**, *9*, 15716. [CrossRef]

80. Liang, C.; Niu, L.; Xiao, Z.; Zheng, C.; Shen, Y.; Shi, Y.; Han, X. Whole-genome sequencing of prostate cancer reveals novel mutation-driven processes and molecular subgroups. *Life Sci.* **2019**, *254*, 117218. [CrossRef]

81. Mendelaar, P.A.J.; Smid, M.; van Riet, J.; Angus, L.; Labots, M.; Steeghs, N.; Hendriks, M.P.; Cirkel, G.A.; van Rooijen, J.M.; Tije, A.J.T.; et al. Whole genome sequencing of metastatic colorectal cancer reveals prior treatment effects and specific metastasis features. *Nat. Commun.* **2021**, *12*, 1–11. [CrossRef]

82. Nair, S.V.; Madhulaxmi; Thomas, G.; Ankathil, R. Next-Generation Sequencing in Cancer. *J. Maxillofac. Oral Surg.* **2020**, *20*, 340–344. [CrossRef] [PubMed]

83. Kamps, R.; Brandão, R.D.; van den Bosch, B.J.; Paulussen, A.D.; Xanthoulea, S.; Blok, M.J.; Romano, A. Next-Generation Sequencing in Oncology: Genetic Diagnosis, Risk Prediction and Cancer Classification. *Int. J. Mol. Sci.* **2017**, *18*, 308. [CrossRef] [PubMed]

84. Meyerson, M.; Gabriel, S.; Getz, G. Advances in understanding cancer genomes through second-generation sequencing. *Nat. Rev. Genet.* **2010**, *11*, 685–696. [CrossRef]

85. Nakagawa, H.; Wardell, C.P.; Furuta, M.; Taniguchi, H.; Fujimoto, A. Cancer whole-genome sequencing: Present and future. *Oncogene* **2015**, *34*, 5943–5950. [CrossRef]

86. Hussen, B.M.; Abdullah, S.T.; Salihi, A.; Sabir, D.K.; Sidiq, K.R.; Rasul, M.F.; Hidayat, H.J.; Ghafouri-Fard, S.; Taheri, M.; Jamali, E. The emerging roles of NGS in clinical oncology and personalized medicine. *Pathol. Res. Pract.* **2022**, *230*, 153760. [CrossRef]

87. Gagan, J.; Van Allen, E.M. Next-generation sequencing to guide cancer therapy. *Genome Med.* **2015**, *7*, 80. [CrossRef]

88. The Cancer Genome Atlas Research Network; Weinstein, J.N.; Collisson, E.A.; Mills, G.B.; Shaw, K.R.M.; Ozenberger, B.A.; Ellrott, K.; Shmulevich, I.; Sander, C.; Stuart, J.M. The Cancer Genome Atlas Pan-Cancer analysis project. *Nat. Genet.* **2013**, *45*, 1113–1120. [CrossRef]

89. NIH. Cancer Genome Atlas. In *Encyclopedia of Genetics, Genomics, Proteomics and Informatics*; Rédei, G.P., Ed.; Springer: Dordrecht, The Netherlands, 2008; p. 265. ISBN 9781402067549.

90. Hudson, T.J.; Anderson, W.; Artez, A.; Barker, A.D.; Bell, C.; Bernabé, R.R.; Bhan, M.K.; Calvo, F.; Eerola, I.; Gerhard, D.S.; et al. International Network of Cancer Genome Projects. *Nature* **2010**, *464*, 993–998. [CrossRef]

91. Trotman, J.; Armstrong, R.; Firth, H.; Trayers, C.; Watkins, J.; Allinson, K.; Jacques, T.S.; Nicholson, J.C.; Burke, G.A.A.; Ambrose, J.C.; et al. The NHS England 100,000 Genomes Project: Feasibility and utility of centralised genome sequencing for children with cancer. *Br. J. Cancer* **2022**, *127*, 137–144. [CrossRef]

92. Turnbull, C. Introducing whole-genome sequencing into routine cancer care: The Genomics England 100,000 Genomes Project. *Ann. Oncol.* **2018**, *29*, 784–787. [CrossRef] [PubMed]

93. Guan, J.; Gupta, R.; Filipp, F.V. Cancer systems biology of TCGA SKCM: Efficient detection of genomic drivers in melanoma. *Sci. Rep.* **2015**, *5*, srep07857. [CrossRef] [PubMed]

94. Wang, M.; Liu, M.; Huang, Y.; Wang, Z.; Wang, Y.; He, K.; Bai, R.; Ying, T.; Zheng, Y. Differential Gene Expression and Methylation Analysis of Melanoma in TCGA Database to Further Study the Expression Pattern of KYNU in Melanoma. *J. Pers. Med.* **2022**, *12*, 1209. [CrossRef] [PubMed]

95. Scatena, C.; Murtas, D.; Tomei, S. Cutaneous Melanoma Classification: The Importance of High-Throughput Genomic Technologies. *Front. Oncol.* **2021**, *11*, 635488. [CrossRef] [PubMed]

96. Ablain, J.; Al Mahi, A.; Rothschild, H.; Prasad, M.; Aires, S.; Yang, S.; Dokukin, M.E.; Xu, S.; Dang, M.; Sokolov, I.; et al. Loss of NECTIN1 triggers melanoma dissemination upon local IGF1 depletion. *Nat. Genet.* **2022**, 1–14. [CrossRef]

97. Scherrer, E.; Hair, G.; Mt-Isa, S.; Pereira, M.; Chan, G.; Shui, I.; Arumugam, P.; Zarowiecki, M.; Witkowska, K.; Rahim, T.; et al. 1136P Feasibility of linking the UK 100,000 genomes project and real-world evidence databases for a melanoma patient population. *Ann. Oncol.* **2020**, *31*, S760–S761. [CrossRef]

98. Griffith, M.; Miller, C.; Griffith, O.; Krysiak, K.; Skidmore, Z.; Ramu, A.; Walker, J.R.; Dang, H.X.; Trani, L.; Larson, D.; et al. Optimizing Cancer Genome Sequencing and Analysis. *Cell Syst.* **2015**, *1*, 210–223. [CrossRef]

99. Ku, C.S.; Cooper, D.N.; Roukos, D.H. Clinical relevance of cancer genome sequencing. *World J. Gastroenterol.* **2013**, *19*, 2011–2018. [CrossRef]

100. Koboldt, D.C. Best practices for variant calling in clinical sequencing. *Genome Med.* **2020**, *12*, 1–13. [CrossRef]

101. Chen, H.; Li, J.; Wang, Y.; Ng, P.K.-S.; Tsang, Y.H.; Shaw, K.R.; Mills, G.; Liang, H. Comprehensive assessment of computational algorithms in predicting cancer driver mutations. *Genome Biol.* **2020**, *21*, 1–17. [CrossRef]

102. Borad, M.J.; Egan, J.B.; Condjella, R.M.; Liang, W.S.; Fonseca, R.; Ritacca, N.R.; McCullough, A.E.; Barrett, M.T.; Hunt, K.S.; Champion, M.D.; et al. Clinical Implementation of Integrated Genomic Profiling in Patients with Advanced Cancers. *Sci. Rep.* **2016**, *6*, 25. [CrossRef] [PubMed]

103. Arora, K.; Shah, M.; Johnson, M.; Sanghvi, R.; Shelton, J.; Nagulapalli, K.; Oschwald, D.M.; Zody, M.C.; Germer, S.; Jobanputra, V.; et al. Deep whole-genome sequencing of 3 cancer cell lines on 2 sequencing platforms. *Sci. Rep.* **2019**, *9*, 1–13. [CrossRef] [PubMed]

104. Garcia, M.; Juhos, S.; Larsson, M.; Olason, P.I.; Martin, M.; Eisfeldt, J.; DiLorenzo, S.; Sandgren, J.; De Ståhl, T.D.; Ewels, P.; et al. Sarek: A portable workflow for whole-genome sequencing analysis of germline and somatic variants. *F1000Research* **2020**, *9*, 63. [CrossRef] [PubMed]

105. Van der Auwera, G.A.; O'Connor, B.D. *Genomics in the Cloud: Using Docker, GATK, and WDL in Terra*; O'Reilly Media, Inc.: Sebastopol, CA, USA, 2020; ISBN 9781491975169.

106. Nakagawa, H.; Fujita, M. Whole genome sequencing analysis for cancer genomics and precision medicine. *Cancer Sci.* **2018**, *109*, 513–522. [CrossRef] [PubMed]

107. Köster, J.; Rahmann, S. Snakemake–a scalable bioinformatics workflow engine. *Bioinformatics* **2012**, *28*, 2520–2522. [CrossRef] [PubMed]

108. Voss, K.; Van der Auwera, G.; Gentry, J. Full-stack genomics pipelining with GATK4 + WDL + Cromwell. *F1000Research* **2017**, *6*. [CrossRef]

109. Di Tommaso, P.; Chatzou, M.; Floden, E.W.; Barja, P.P.; Palumbo, E.; Notredame, C. Nextflow enables reproducible computational workflows. *Nat. Biotechnol.* **2017**, *35*, 316–319. [CrossRef]

110. Li, H.; Dawood, M.; Khayat, M.M.; Farek, J.R.; Jhangiani, S.N.; Khan, Z.M.; Mitani, T.; Coban-Akdemir, Z.; Lupski, J.R.; Venner, E.; et al. Exome variant discrepancies due to reference-genome differences. *Am. J. Hum. Genet.* **2021**, *108*, 1239–1250. [CrossRef]

111. Nurk, S.; Koren, S.; Rhie, A.; Rautiainen, M.; Bzikadze, A.V.; Mikheenko, A.; Vollger, M.R.; Altemose, N.; Uralsky, L.; Gershman, A.; et al. The complete sequence of a human genome. *Science* **2022**, *376*, 44–53. [CrossRef]

112. Aganezov, S.; Yan, S.M.; Soto, D.C.; Kirsche, M.; Zarate, S.; Avdeyev, P.; Taylor, D.J.; Shafin, K.; Shumate, A.; Xiao, C.; et al. A complete reference genome improves analysis of human genetic variation. *Science* **2022**, *376*, eabl3533. [CrossRef]

113. Alkan, C.; Carbone, L.; Dennis, M.; Ernst, J.; Evrony, G.; Girirajan, S.; Leung, D.C.Y.; Cheng, C.C.; MacAlpine, D.; Ni, T.; et al. Implications of the first complete human genome assembly. *Genome Res.* **2022**, *32*, 595–598. [CrossRef]

114. Li, H. Aligning Sequence Reads, Clone Sequences and Assembly Contigs with BWA-MEM. *arXiv* **2013**, arXiv:1303.3997.

115. Mu, J.C.; Jiang, H.; Kiani, A.; Mohiyuddin, M.; Asadi, N.B.; Wong, W.H. Fast and accurate read alignment for resequencing. *Bioinformatics* **2012**, *28*, 2366–2373. [CrossRef] [PubMed]
116. Langmead, B.; Salzberg, S.L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **2012**, *9*, 357–359. [CrossRef] [PubMed]
117. Li, H. Minimap2: Pairwise alignment for nucleotide sequences. *Bioinformatics* **2018**, *34*, 3094–3100. [CrossRef] [PubMed]
118. Danecek, P.; Bonfield, J.K.; Liddle, J.; Marshall, J.; Ohan, V.; Pollard, M.O.; Whitwham, A.; Keane, T.; McCarthy, S.A.; Davies, R.M.; et al. Twelve years of SAMtools and BCFtools. *GigaScience* **2021**, *10*, giab008. [CrossRef] [PubMed]
119. Tarasov, A.; Vilella, A.J.; Cuppen, E.; Nijman, I.J.; Prins, P. Sambamba: Fast processing of NGS alignment formats. *Bioinformatics* **2015**, *31*, 2032–2034. [CrossRef]
120. Broad Institute. Picard Toolkit. Available online: http://broadinstitute.github.io/picard/ (accessed on 10 October 2022).
121. Pedersen, B.S.; Quinlan, A.R. Mosdepth: Quick Coverage Calculation for Genomes and Exomes. *Bioinformatics* **2018**, *34*, 867–868. [CrossRef]
122. Okonechnikov, K.; Conesa, A.; García-Alcalde, F. Qualimap 2: Advanced multi-sample quality control for high-throughput sequencing data. *Bioinformatics* **2016**, *32*, 292–294. [CrossRef]
123. Ewels, P.; Magnusson, M.; Lundin, S.; Käller, M. MultiQC: Summarize analysis results for multiple tools and samples in a single report. *Bioinformatics* **2016**, *32*, 3047–3048. [CrossRef]
124. Benjamin, D.; Sato, T.; Cibulskis, K.; Getz, G.; Stewart, C.; Lichtenstein, L. Calling Somatic SNVs and Indels with Mutect2. *BioRxiv* **2019**, 861054. [CrossRef]
125. Kim, S.; Scheffler, K.; Halpern, A.L.; Bekritsky, M.A.; Noh, E.; Källberg, M.; Chen, X.; Kim, Y.; Beyter, D.; Krusche, P.; et al. Strelka2: Fast and accurate calling of germline and somatic variants. *Nat. Methods* **2018**, *15*, 591–594. [CrossRef]
126. Koboldt, D.C.; Zhang, Q.; Larson, D.E.; Shen, D.; McLellan, M.D.; Lin, L.; Miller, C.A.; Mardis, E.R.; Ding, L.; Wilson, R.K. VarScan 2: Somatic mutation and copy number alteration discovery in cancer by exome sequencing. *Genome Res.* **2012**, *22*, 568–576. [CrossRef]
127. Poplin, R.; Chang, P.-C.; Alexander, D.; Schwartz, S.; Colthurst, T.; Ku, A.; Newburger, D.; Dijamco, J.; Nguyen, N.; Afshar, P.T.; et al. A universal SNP and small-indel variant caller using deep neural networks. *Nat. Biotechnol.* **2018**, *36*, 983–987. [CrossRef] [PubMed]
128. Patiyal, S.; Dhall, A.; Raghava, G.P.S. Prediction of risk-associated genes and high-risk liver cancer patients from their mutation profile: Benchmarking of mutation calling techniques. *Biol. Methods Protoc.* **2022**, *7*, bpac012. [CrossRef]
129. Zhou, H.; Hu, Y.; Luo, R.; Zhao, Y.; Pan, H.; Ji, L.; Zhou, T.; Zhang, L.; Long, H.; Fu, J.; et al. Multi-region exome sequencing reveals the intratumoral heterogeneity of surgically resected small cell lung cancer. *Nat. Commun.* **2021**, *12*, 1–11. [CrossRef]
130. Ura, H.; Togi, S.; Niida, Y. Dual Deep Sequencing Improves the Accuracy of Low-Frequency Somatic Mutation Detection in Cancer Gene Panel Testing. *Int. J. Mol. Sci.* **2020**, *21*, 3530. [CrossRef] [PubMed]
131. Lai, Z.; Markovets, A.; Ahdesmäki, M.; Chapman, B.; Hofmann, O.; McEwen, R.; Johnson, J.; Dougherty, B.; Barrett, J.C.; Dry, J.R. VarDict: A novel and versatile variant caller for next-generation sequencing in cancer research. *Nucleic Acids Res.* **2016**, *44*, e108. [CrossRef] [PubMed]
132. Lai, Z.; Brosnan, M.; Sokol, E.S.; Xie, M.; Dry, J.R.; Harrington, E.A.; Barrett, J.C.; Hodgson, D. Landscape of homologous recombination deficiencies in solid tumours: Analyses of two independent genomic datasets. *BMC Cancer* **2022**, *22*, 1–13. [CrossRef] [PubMed]
133. Talevich, E.; Shain, A.H.; Botton, T.; Bastian, B.C. CNVkit: Genome-Wide Copy Number Detection and Visualization from Targeted DNA Sequencing. *PLoS Comput. Biol.* **2015**, *12*, e1004873. [CrossRef]
134. Vittoria, M.A.; Kingston, N.; Kotynkova, K.; Xia, E.; Hong, R.; Huang, L.; McDonald, S.; Tilston-Lunel, A.; Darp, R.; Campbell, J.D.; et al. Inactivation of the Hippo tumor suppressor pathway promotes melanoma. *Nat. Commun.* **2022**, *13*, 1–17. [CrossRef] [PubMed]
135. Chen, X.; Schulz-Trieglaff, O.; Shaw, R.; Barnes, B.; Schlesinger, F.; Källberg, M.; Cox, A.J.; Kruglyak, S.; Saunders, C.T. Manta: Rapid detection of structural variants and indels for germline and cancer sequencing applications. *Bioinformatics* **2016**, *32*, 1220–1222. [CrossRef] [PubMed]
136. Yu, Y.; Zhang, Z.; Dong, X.; Yang, R.; Duan, Z.; Xiang, Z.; Li, J.; Li, G.; Yan, F.; Xue, H.; et al. Pangenomic analysis of Chinese gastric cancer. *Nat. Commun.* **2022**, *13*, 1–13. [CrossRef] [PubMed]
137. Rausch, T.; Zichner, T.; Schlattl, A.; Stütz, A.M.; Benes, V.; Korbel, J. DELLY: Structural variant discovery by integrated paired-end and split-read analysis. *Bioinformatics* **2012**, *28*, i333–i339. [CrossRef] [PubMed]
138. Seo, J.; Kim, H.; Min, K.I.; Kim, C.; Kwon, Y.; Zheng, Z.; Kim, Y.; Park, H.-S.; Ju, Y.S.; Roh, M.R.; et al. Weight-bearing activity impairs nuclear membrane and genome integrity via YAP activation in plantar melanoma. *Nat. Commun.* **2022**, *13*, 1–15. [CrossRef]
139. Layer, R.M.; Chiang, C.; Quinlan, A.R.; Hall, I.M. LUMPY: A probabilistic framework for structural variant discovery. *Genome Biol.* **2014**, *15*, R84. [CrossRef]
140. Akdemir, K.C.; Le, V.T.; Chandran, S.; Li, Y.; Verhaak, R.G.; Beroukhim, R.; Campbell, P.J.; Chin, L.; Dixon, J.R.; Futreal, P.A.; et al. Disruption of chromatin folding domains by somatic genomic rearrangements in human cancer. *Nat. Genet.* **2020**, *52*, 294–305. [CrossRef]

141. Cameron, D.L.; Baber, J.; Shale, C.; Valle-Inclan, J.E.; Besselink, N.; van Hoeck, A.; Janssen, R.; Cuppen, E.; Priestley, P.; Papenfuss, A.T. GRIDSS2: Comprehensive characterisation of somatic structural variation using single breakend variants and structural variant phasing. *Genome Biol.* **2021**, *22*, 1–25. [CrossRef]

142. Tiong, M.I.S.; Wilson, B.C.; Yerneni, M.S.; Markham, J.; Dun, B.K.; Bajel, F.A.; Thompson, B.E.R.; Westerman, M.D.A.; Blombery, M.P. Mutational and Copy Number Profiling of Circulating Tumor DNA in Acute Myeloid Leukemia Using Targeted Next Generation Sequencing. *Blood* **2020**, *136*, 39–40. [CrossRef]

143. Field, M.G.; Durante, M.A.; Anbunathan, H.; Cai, L.Z.; Decatur, C.L.; Bowcock, A.M.; Kurtenbach, S.; Harbour, J.W. Punctuated evolution of canonical genomic aberrations in uveal melanoma. *Nat. Commun.* **2018**, *9*, 116. [CrossRef]

144. Demidov, G.; Ossowski, S. ClinCNV: Novel Method for Allele-Specific Somatic Copy-Number Alterations Detection. *bioRxiv* **2019**, 837971. [CrossRef]

145. Prasad, A.; Rabionet, R.; Espinet, B.; Zapata, L.; Puiggros, A.; Melero, C.; Puig, A.; Sarria-Trujillo, Y.; Ossowski, S.; Garcia-Muret, M.P.; et al. Identification of Gene Mutations and Fusion Genes in Patients with Sézary Syndrome. *J. Investig. Dermatol.* **2016**, *136*, 1490–1499. [CrossRef]

146. Plagnol, V.; Curtis, J.; Epstein, M.; Mok, K.; Stebbings, E.; Grigoriadou, S.; Wood, N.; Hambleton, S.; Burns, S.; Thrasher, A.; et al. A robust model for read count data in exome sequencing experiments and implications for copy number variant calling. *Bioinformatics* **2012**, *28*, 2747–2754. [CrossRef]

147. Boujemaa, M.; Hamdi, Y.; Mejri, N.; Romdhane, L.; Ghedira, K.; Bouaziz, H.; El Benna, H.; Labidi, S.; Dallali, H.; Jaidane, O.; et al. Germline copy number variations in BRCA1/2 negative families: Role in the molecular etiology of hereditary breast cancer in Tunisia. *PLoS ONE* **2021**, *16*, e0245362. [CrossRef]

148. Minoche, A.E.; Ben Lundie, B.; Peters, G.B.; Ohnesorg, T.; Pinese, M.; Thomas, D.M.; Zankl, A.; Roscioli, T.; Schonrock, N.; Kummerfeld, S.; et al. ClinSV: Clinical grade structural and copy number variant detection from whole genome sequencing data. *Genome Med.* **2021**, *13*, 1–19. [CrossRef]

149. Deng, N.; Minoche, A.; Harvey, K.; Li, M.; Winkler, J.; Goga, A.; Swarbrick, A. Deep whole genome sequencing identifies recurrent genomic alterations in commonly used breast cancer cell lines and patient-derived xenograft models. *Breast Cancer Res.* **2022**, *24*, 1–12. [CrossRef] [PubMed]

150. Garcia-Prieto, C.A.; Martínez-Jiménez, F.; Valencia, A.; Porta-Pardo, E. Detection of oncogenic and clinically actionable mutations in cancer genomes critically depends on variant calling tools. *Bioinformatics* **2022**, *38*, 3181–3191. [CrossRef]

151. Xu, C. A review of somatic single nucleotide variant calling algorithms for next-generation sequencing data. *Comput. Struct. Biotechnol. J.* **2018**, *16*, 15–24. [CrossRef]

152. Danecek, P.; Auton, A.; Abecasis, G.; Albers, C.A.; Banks, E.; DePristo, M.A.; Handsaker, R.E.; Lunter, G.; Marth, G.T.; Sherry, S.T.; et al. The variant call format and VCFtools. *Bioinformatics* **2011**, *27*, 2156–2158. [CrossRef] [PubMed]

153. Shao, X.; Lv, N.; Liao, J.; Long, J.; Xue, R.; Ai, N.; Xu, D.; Fan, X. Copy number variation is highly correlated with differential gene expression: A pan-cancer study. *BMC Med. Genet.* **2019**, *20*, 1–14. [CrossRef]

154. Cho, S. Set-Wise Differential Interaction Between Copy Number Alterations and Gene Expressions of Lower-Grade Glioma Reveals Prognosis-Associated Pathways. *Entropy* **2020**, *22*, 1434. [CrossRef]

155. Shahrisa, A.; Tahmasebi-Birgani, M.; Ansari, H.; Mohammadi, Z.; Carloni, V.; Asl, J.M. The pattern of gene copy number alteration (CNAs) in hepatocellular carcinoma: An in silico analysis. *Mol. Cytogenet.* **2021**, *14*, 1–10. [CrossRef]

156. van Belzen, I.A.E.M.; Schönhuth, A.; Kemmeren, P.; Hehir-Kwa, J.Y. Structural variant detection in cancer genomes: Computational challenges and perspectives for precision oncology. *Npj Precis. Oncol.* **2021**, *5*, 1–11. [CrossRef]

157. Gong, T.; Hayes, V.M.; Chan, E.K.F. Detection of somatic structural variants from short-read next-generation sequencing data. *Brief. Bioinform.* **2020**, *22*, bbaa056. [CrossRef]

158. Cameron, D.L.; Di Stefano, L.; Papenfuss, A.T. Comprehensive evaluation and characterisation of short read general-purpose structural variant calling software. *Nat. Commun.* **2019**, *10*, 1–11. [CrossRef]

159. Coutelier, M.; Holtgrewe, M.; Jäger, M.; Flöttman, R.; Mensah, M.A.; Spielmann, M.; Krawitz, P.; Horn, D.; Beule, D.; Mundlos, S. Combining callers improves the detection of copy number variants from whole-genome sequencing. *Eur. J. Hum. Genet.* **2021**, *30*, 178–186. [CrossRef]

160. Olson, N.D.; Wagner, J.; McDaniel, J.; Stephens, S.H.; Westreich, S.T.; Prasanna, A.G.; Johanson, E.; Boja, E.; Maier, E.J.; Serang, O.; et al. PrecisionFDA Truth Challenge V2: Calling variants from short and long reads in difficult-to-map regions. *Cell Genom.* **2022**, *2*, 100129. [CrossRef]

161. Wagner, J.; Olson, N.D.; Harris, L.; McDaniel, J.; Cheng, H.; Fungtammasan, A.; Hwang, Y.-C.; Gupta, R.; Wenger, A.M.; Rowell, W.J.; et al. Curated variation benchmarks for challenging medically relevant autosomal genes. *Nat. Biotechnol.* **2022**, *40*, 672–680. [CrossRef]

162. Li, M.M.; Datto, M.; Duncavage, E.J.; Kulkarni, S.; Lindeman, N.I.; Roy, S.; Tsimberidou, A.M.; Vnencak-Jones, C.L.; Wolff, D.J.; Younes, A.; et al. Standards and Guidelines for the Interpretation and Reporting of Sequence Variants in Cancer: A Joint Consensus Recommendation of the Association for Molecular Pathology, American Society of Clinical Oncology, and College of American Pathologists. *J. Mol. Diagn.* **2017**, *19*, 4–23. [CrossRef]

163. Sherry, S.T.; Ward, M.-H.; Kholodov, M.; Baker, J.; Phan, L.; Smigielski, E.M.; Sirotkin, K. dbSNP: The NCBI database of genetic variation. *Nucleic Acids Res.* **2001**, *29*, 308–311. [CrossRef]

164. Karczewski, K.J.; Francioli, L.C.; Tiao, G.; Cummings, B.B.; Alfoldi, J.; Wang, Q.; Collins, R.L.; Laricchia, K.M.; Ganna, A.; Birnbaum, D.P.; et al. The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature* **2020**, *581*, 434–443. [CrossRef]

165. Tate, J.G.; Bamford, S.; Jubb, H.C.; Sondka, Z.; Beare, D.M.; Bindal, N.; Boutselakis, H.; Cole, C.G.; Creatore, C.; Dawson, E.; et al. COSMIC: The Catalogue of Somatic Mutations in Cancer. *Nucleic Acids Res.* **2019**, *47*, D941–D947. [CrossRef]

166. McLaren, W.; Gil, L.; Hunt, S.E.; Riat, H.S.; Ritchie, G.R.S.; Thormann, A.; Flicek, P.; Cunningham, F. The Ensembl Variant Effect Predictor. *Genome Biol.* **2016**, *17*, 1–14. [CrossRef]

167. Wang, K.; Li, M.; Hakonarson, H. ANNOVAR: Functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.* **2010**, *38*, e164. [CrossRef]

168. DePristo, M.A.; Banks, E.; Poplin, R.; Garimella, K.V.; Maguire, J.R.; Hartl, C.; Philippakis, A.A.; Del Angel, G.; Rivas, M.A.; Hanna, M.; et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat. Genet.* **2011**, *43*, 491–498. [CrossRef]

169. Dashti, M.J.S.; Gamieldien, J. A practical guide to filtering and prioritizing genetic variants. *BioTechniques* **2017**, *62*, 18–30. [CrossRef]

170. Thorvaldsdóttir, H.; Robinson, J.T.; Mesirov, J.P. Integrative Genomics Viewer (IGV): High-performance genomics data visualization and exploration. *Brief. Bioinform.* **2013**, *14*, 178–192. [CrossRef]

171. Belyeu, J.R.; Chowdhury, M.; Brown, J.; Pedersen, B.S.; Cormier, M.J.; Quinlan, A.R.; Layer, R.M. Samplot: A platform for structural variant visual validation and automated filtering. *Genome Biol.* **2021**, *22*, 1–13. [CrossRef]

172. Liu, Z.; Roberts, R.; Mercer, T.R.; Xu, J.; Sedlazeck, F.J.; Tong, W. Towards accurate and reliable resolution of structural variants for clinical diagnosis. *Genome Biol.* **2022**, *23*, 1–25. [CrossRef]

173. Griffith, M.; Spies, N.C.; Krysiak, K.; McMichael, J.F.; Coffman, A.C.; Danos, A.M.; Ainscough, B.J.; Ramirez, C.A.; Rieke, D.T.; Kujan, L.; et al. CIViC is a community knowledgebase for expert crowdsourcing the clinical interpretation of variants in cancer. *Nat. Genet.* **2017**, *49*, 170–174. [CrossRef]

174. Ainscough, B.J.; Barnell, E.K.; Ronning, P.; Campbell, K.M.; Wagner, A.H.; Fehniger, T.A.; Dunn, G.P.; Uppaluri, R.; Govindan, R.; Rohan, T.E.; et al. A deep learning approach to automate refinement of somatic variant calling from cancer sequencing data. *Nat. Genet.* **2018**, *50*, 1735–1743. [CrossRef]

175. Vaisband, M.; Schubert, M.; Gassner, F.J.; Geisberger, R.; Greil, R.; Zaborsky, N.; Hasenauer, J. Validation of Genetic Variants from NGS Data Using Deep Convolutional Neural Networks. *bioRxiv* **2022**, 488021. [CrossRef]

176. Myers, M.A.; Zaccaria, S.; Raphael, B.J. Identifying tumor clones in sparse single-cell mutation data. *Bioinformatics* **2020**, *36*, i186–i193. [CrossRef]

177. Zhang, X.; Lv, D.; Zhang, Y.; Liu, Q.; Li, Z. Clonal evolution of acute myeloid leukemia highlighted by latest genome sequencing studies. *Oncotarget* **2016**, *7*, 58586–58594. [CrossRef]

178. Strom, S.P. Current practices and guidelines for clinical next-generation sequencing oncology testing. *Cancer Biol. Med.* **2016**, *13*, 3–11. [CrossRef]

179. Ding, L.; Ley, T.J.; Larson, D.E.; Miller, C.A.; Koboldt, D.C.; Welch, J.S.; Ritchey, J.K.; Young, M.A.; Lamprecht, T.; McLellan, M.D.; et al. Clonal evolution in relapsed acute myeloid leukaemia revealed by whole-genome sequencing. *Nature* **2012**, *481*, 506–510. [CrossRef]

180. Gatalica, Z.; Xiu, J.; Swensen, J.; Vranic, S. Molecular characterization of cancers with NTRK gene fusions. *Mod. Pathol.* **2018**, *32*, 147–153. [CrossRef]

181. Quan, V.L.; Panah, E.; Zhang, B.; Shi, K.; Mohan, L.S.; Gerami, P. The role of gene fusions in melanocytic neoplasms. *J. Cutan. Pathol.* **2019**, *46*, 878–887. [CrossRef]

182. Chen, H.-F.; Wang, W.-X.; Xu, C.-W.; Huang, L.-C.; Li, X.-F.; Lan, G.; Zhai, Z.-Q.; Zhu, Y.-C.; Du, K.-Q.; Lei, L.; et al. A novel SOS1-ALK fusion variant in a patient with metastatic lung adenocarcinoma and a remarkable response to crizotinib. *Lung Cancer* **2020**, *142*, 59–62. [CrossRef]

183. Mittal, V.K.; McDonald, J.F. De novo assembly and characterization of breast cancer transcriptomes identifies large numbers of novel fusion-gene transcripts of potential functional significance. *BMC Med. Genom.* **2017**, *10*, 1–20. [CrossRef]

184. Bolger, A.M.; Lohse, M.; Usadel, B. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* **2014**, *30*, 2114–2120. [CrossRef]

185. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet. J.* **2011**, *17*, 10–12. [CrossRef]

186. Krueger, F.; James, F.; Ewels, P.; Afyounian, E.; Schuster-Boeckler, B. TrimGalore: V0.6.7—DOI via Zenodo. 2021. Available online: https://zenodo.org/record/5127899 (accessed on 15 October 2022).

187. Xie, Y.; Wu, G.; Tang, J.; Luo, R.; Patterson, J.; Liu, S.; Huang, W.; He, G.; Gu, S.; Li, S.; et al. SOAPdenovo-Trans: De novo transcriptome assembly with short RNA-Seq reads. *Bioinformatics* **2014**, *30*, 1660–1666. [CrossRef]

188. Grabherr, M.G.; Haas, B.J.; Yassour, M.; Levin, J.Z.; Thompson, D.A.; Amit, I.; Adiconis, X.; Fan, L.; Raychowdhury, R.; Zeng, Q.D.; et al. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotechnol.* **2011**, *29*, 644–652. [CrossRef]

189. Robertson, G.; Schein, J.; Chiu, R.; Corbett, R.; Field, M.; Jackman, S.D.; Mungall, K.; Lee, S.; Okada, H.M.; Qian, J.Q.; et al. De novo assembly and analysis of RNA-seq data. *Nat. Methods* **2010**, *7*, 909–912. [CrossRef]

190. Chiu, R.; Nip, K.M.; Chu, J.; Birol, I. TAP: A targeted clinical genomics pipeline for detecting transcript variants using RNA-seq data. *BMC Med. Genom.* **2018**, *11*, 79. [CrossRef]

191. Raghavan, V.; Kraft, L.; Mesny, F.; Rigerte, L. A simple guide to de novo transcriptome assembly and annotation. *Brief. Bioinform.* **2022**, *23*, bbab563. [CrossRef]

192. Haas, B.J.; Papanicolaou, A.; Yassour, M.; Grabherr, M.; Blood, P.D.; Bowden, J.; Couger, M.B.; Eccles, D.; Li, B.; Lieber, M.; et al. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat. Protoc.* **2013**, *8*, 1494–1512. [CrossRef]

193. Chiu, R.; Nip, K.M.; Birol, I. Fusion-Bloom: Fusion detection in assembled transcriptomes. *Bioinformatics* **2019**, *36*, 2256–2257. [CrossRef]

194. Zhang, J.; White, N.M.; Schmidt, H.K.; Fulton, R.S.; Tomlinson, C.; Warren, W.C.; Wilson, R.K.; Maher, C.A. INTEGRATE: Gene fusion discovery using whole genome and transcriptome data. *Genome Res.* **2015**, *26*, 108–118. [CrossRef]

195. Endrullat, C.; Glökler, J.; Franke, P.; Frohme, M. Standardization and quality management in next-generation sequencing. *Appl. Transl. Genom.* **2016**, *10*, 2–9. [CrossRef]

196. Andrews, S. FastQC: A Quality Control Tool for High Throughput Sequence Data 2010. Available online: https://www.scienceopen.com/document?vid=de674375-ab83-4595-afa9-4c8aa9e4e736 (accessed on 15 October 2022).

197. Manichaikul, A.; Mychaleckyj, J.C.; Rich, S.S.; Daly, K.; Sale, M.; Chen, W.-M. Robust relationship inference in genome-wide association studies. *Bioinformatics* **2010**, *26*, 2867–2873. [CrossRef]

198. Thornton, T.; Tang, H.; Hoffmann, T.J.; Ochs-Balcom, H.M.; Caan, B.J.; Risch, N. Estimating Kinship in Admixed Populations. *Am. J. Hum. Genet.* **2012**, *91*, 122–138. [CrossRef]

199. Lee, H.; Chen, L. Inference of kinship using spatial distributions of SNPs for genome-wide association studies. *BMC Genom.* **2016**, *17*, 1–9. [CrossRef]

200. Pedersen, B.S.; Bhetariya, P.J.; Brown, J.; Kravitz, S.N.; Marth, G.; Jensen, R.L.; Bronner, M.P.; Underhill, H.R.; Quinlan, A.R. Somalier: Rapid relatedness estimation for cancer and germline studies using efficient genome sketches. *Genome Med.* **2020**, *12*, 1–9. [CrossRef]

201. Webster, T.H.; Couse, M.; Grande, B.M.; Karlins, E.; Phung, T.N.; Richmond, P.A.; Whitford, W.; Wilson Sayres, M.A. Identifying, Understanding, and Correcting Technical Biases on the Sex Chromosomes in next-Generation Sequencing Data. *bioRxiv* **2018**, 346940. [CrossRef]

202. Genomics Division, ITER SexQC-for-NGS-Data: Sex Quality Control for Next Generation Sequencing Data; Github. Available online: https://github.com/genomicsITER/sexQC-for-NGS-data (accessed on 15 October 2022).

203. Pollard, M.O.; Gurdasani, D.; Mentzer, A.J.; Porter, T.; Sandhu, M.S. Long reads: Their purpose and place. *Hum. Mol. Genet.* **2018**, *27*, R234–R241. [CrossRef]

204. Sakamoto, Y.; Sereewattanawoot, S.; Suzuki, A. A new era of long-read sequencing for cancer genomics. *J. Hum. Genet.* **2019**, *65*, 3–10. [CrossRef]

205. Xia, L.C.; Bell, J.M.; Wood-Bouwens, C.; Chen, J.J.; Zhang, N.R.; Ji, H.P. Identification of large rearrangements in cancer genomes with barcode linked reads. *Nucleic Acids Res.* **2017**, *46*, e19. [CrossRef]

206. Dozmorov, M.G.; Tyc, K.M.; Sheffield, N.C.; Boyd, D.C.; Olex, A.L.; Reed, J.; Harrell, J.C. Chromatin conformation capture (Hi-C) sequencing of patient-derived xenografts: Analysis guidelines. *GigaScience* **2021**, *10*, giab022. [CrossRef]

207. Chan, E.K.; Cameron, D.L.; Petersen, D.C.; Lyons, R.J.; Baldi, B.F.; Papenfuss, A.T.; Thomas, D.M.; Hayes, V.M. Optical mapping reveals a higher level of genomic architecture of chained fusions in cancer. *Genome Res.* **2018**, *28*, 726–738. [CrossRef]

208. Euskirchen, P.; Bielle, F.; Labreche, K.; Kloosterman, W.P.; Rosenberg, S.; Daniau, M.; Schmitt, C.; Masliah-Planchon, J.; Bourdeaut, F.; Dehais, C.; et al. Same-day genomic and epigenomic diagnosis of brain tumors using real-time nanopore sequencing. *Acta Neuropathol.* **2017**, *134*, 691–703. [CrossRef]

209. Sakamoto, Y.; Miyake, S.; Oka, M.; Kanai, A.; Kawai, Y.; Nagasawa, S.; Shiraishi, Y.; Tokunaga, K.; Kohno, T.; Seki, M.; et al. Phasing analysis of lung cancer genomes using a long read sequencer. *Nat. Commun.* **2022**, *13*, 1–17. [CrossRef]

210. Viswanathan, S.R.; Ha, G.; Hoff, A.M.; Wala, J.A.; Carrot-Zhang, J.; Whelan, C.W.; Haradhvala, N.J.; Freeman, S.S.; Reed, S.C.; Rhoades, J.; et al. Structural Alterations Driving Castration-Resistant Prostate Cancer Revealed by Linked-Read Genome Sequencing. *Cell* **2018**, *174*, 433–447.e19. [CrossRef]

211. Greer, S.U.; Nadauld, L.D.; Lau, B.T.; Chen, J.; Wood-Bouwens, C.; Ford, J.M.; Kuo, C.J.; Ji, H.P. Linked read sequencing resolves complex genomic rearrangements in gastric cancer metastases. *Genome Med.* **2017**, *9*, 1–17. [CrossRef]

212. Ren, B.; Yang, J.; Wang, C.; Yang, G.; Wang, H.; Chen, Y.; Xu, R.; Fan, X.; You, L.; Zhang, T.; et al. High-resolution Hi-C maps highlight multiscale 3D epigenome reprogramming during pancreatic cancer metastasis. *J. Hematol. Oncol.* **2021**, *14*, 1–19. [CrossRef]

213. Suttorp, J.; Lühmann, J.L.; Behrens, Y.L.; Göhring, G.; Steinemann, D.; Reinhardt, D.; von Neuhoff, N.; Schneider, M. Optical Genome Mapping as a Diagnostic Tool in Pediatric Acute Myeloid Leukemia. *Cancers* **2022**, *14*, 2058. [CrossRef]

214. Magi, A.; Semeraro, R.; Mingrino, A.; Giusti, B.; D'Aurizio, R. Nanopore sequencing data analysis: State of the art, applications and challenges. *Brief. Bioinform.* **2017**, *19*, 1256–1272. [CrossRef]

215. Jain, M.; Olsen, H.E.; Paten, B.; Akeson, M. The Oxford Nanopore MinION: Delivery of nanopore sequencing to the genomics community. *Genome Biol.* **2016**, *17*, 1–11. [CrossRef]

216. Lu, H.; Giordano, F.; Ning, Z. Oxford Nanopore MinION Sequencing and Genome Assembly. *Genom. Proteom. Bioinform.* **2016**, *14*, 265–279. [CrossRef]

217. Wang, Y.; Zhao, Y.; Bollas, A.; Wang, Y.; Au, K.F. Nanopore sequencing technology, bioinformatics and applications. *Nat. Biotechnol.* **2021**, *39*, 1348–1365. [CrossRef]

218. Rang, F.J.; Kloosterman, W.P.; De Ridder, J. From squiggle to basepair: Computational approaches for improving nanopore sequencing read accuracy. *Genome Biol.* **2018**, *19*, 1–11. [CrossRef]

219. Rhoads, A.; Au, K.F. PacBio Sequencing and Its Applications. *Genom. Proteom. Bioinform.* **2015**, *13*, 278–289. [CrossRef]

220. Wenger, A.M.; Peluso, P.; Rowell, W.J.; Chang, P.-C.; Hall, R.J.; Concepcion, G.T.; Ebler, J.; Fungtammasan, A.; Kolesnikov, A.; Olson, N.D.; et al. Accurate circular consensus long-read sequencing improves variant detection and assembly of a human genome. *Nat. Biotechnol.* **2019**, *37*, 1155–1162. [CrossRef]

221. Ott, A.; Schnable, J.C.; Yeh, C.T.; Wu, L.; Liu, C.; Hu, H.C.; Dalgard, C.L.; Sarkar, S.; Schnable, P.S. Linked read technology for assembling large complex and polyploid genomes. *BMC Genom.* **2018**, *19*, 651. [CrossRef]

222. Van Berkum, N.L.; Lieberman-Aiden, E.; Williams, L.; Imakaev, M.; Gnirke, A.; Mirny, L.A.; Dekker, J.; Lander, E.S. Hi-C: A Method to Study the Three-dimensional Architecture of Genomes. *J. Vis. Exp.* **2010**, *39*, e1869. [CrossRef]

223. Yardımcı, G.G.; Ozadam, H.; Sauria, M.E.G.; Ursu, O.; Yan, K.-K.; Yang, T.; Chakraborty, A.; Kaul, A.; Lajoie, B.R.; Song, F.; et al. Measuring the reproducibility and quality of Hi-C data. *Genome Biol.* **2019**, *20*, 1–19. [CrossRef]

224. Kyriakidou, M.; Tai, H.H.; Anglin, N.L.; Ellis, D.; Strömvik, M.V. Current Strategies of Polyploid Plant Genome Sequence Assembly. *Front. Plant Sci.* **2018**, *9*, 1660. [CrossRef]

225. Oluwadare, O.; Highsmith, M.; Cheng, J. An Overview of Methods for Reconstructing 3-D Chromosome and Genome Structures from Hi-C Data. *Biol. Proced. Online* **2019**, *21*, 7. [CrossRef]

226. Shi, L.; Guo, Y.; Dong, C.; Huddleston, J.; Yang, H.; Han, X.; Fu, A.; Li, Q.; Li, N.; Gong, S.; et al. Long-read sequencing and de novo assembly of a Chinese genome. *Nat. Commun.* **2016**, *7*, 12065. [CrossRef]

227. Sahajpal, N.S.; Lai, C.-Y.J.; Hastie, A.; Mondal, A.K.; Dehkordi, S.R.; van der Made, C.I.; Fedrigo, O.; Al-Ajli, F.; Jalnapurkar, S.; Byrska-Bishop, M.; et al. Optical genome mapping identifies rare structural variations as predisposition factors associated with severe COVID-19. *iScience* **2022**, *25*, 103760. [CrossRef]

228. Goldrich, D.; LaBarge, B.; Chartrand, S.; Zhang, L.; Sadowski, H.; Zhang, Y.; Pham, K.; Way, H.; Lai, C.-Y.; Pang, A.; et al. Identification of Somatic Structural Variants in Solid Tumors by Optical Genome Mapping. *J. Pers. Med.* **2021**, *11*, 142. [CrossRef]

229. Miga, K.H.; Koren, S.; Rhie, A.; Vollger, M.R.; Gershman, A.; Bzikadze, A.; Brooks, S.; Howe, E.; Porubsky, D.; Logsdon, G.A.; et al. Telomere-to-telomere assembly of a complete human X chromosome. *Nature* **2020**, *585*, 79–84. [CrossRef]

230. Jain, M.; Koren, S.; Miga, K.H.; Quick, J.; Rand, A.C.; Sasani, T.A.; Tyson, J.R.; Beggs, A.D.; Dilthey, A.T.; Fiddes, I.T.; et al. Nanopore sequencing and assembly of a human genome with ultra-long reads. *Nat. Biotechnol.* **2018**, *36*, 338–345. [CrossRef]

231. Ardui, S.; Ameur, A.; Vermeesch, J.R.; Hestand, M.S. Single molecule real-time (SMRT) sequencing comes of age: Applications and utilities for medical diagnostics. *Nucleic Acids Res.* **2018**, *46*, 2159–2168. [CrossRef] [PubMed]

232. Tan, K.-T.; Slevin, M.K.; Meyerson, M.; Li, H. Identifying and correcting repeat-calling errors in nanopore sequencing of telomeres. *Genome Biol.* **2022**, *23*, 1–16. [CrossRef]

233. Jones, A.; Torkel, C.; Stanley, D.; Nasim, J.; Borevitz, J.; Schwessinger, B. High-molecular weight DNA extraction, clean-up and size selection for long-read sequencing. *PLoS ONE* **2021**, *16*, e0253830. [CrossRef]

234. Logsdon, G.A.; Vollger, M.R.; Eichler, E.E. Long-read human genome sequencing and its applications. *Nat. Rev. Genet.* **2020**, *21*, 597–614. [CrossRef]

235. Shumate, A.; Zimin, A.V.; Sherman, R.M.; Puiu, D.; Wagner, J.M.; Olson, N.D.; Pertea, M.; Salit, M.L.; Zook, J.M.; Salzberg, S.L. Assembly and annotation of an Ashkenazi human reference genome. *Genome Biol.* **2020**, *21*, 1–18. [CrossRef]

236. Xiao, C.; Chen, Z.; Chen, W.; Padilla, C.; Fang, L.-T.; Liu, T.; Schneider, V.; Wang, C.; Xiao, W. Personalized Genome Assembly for Accurate Cancer Somatic Mutation Discovery Using Cancer-Normal Paired Reference Samples. *bioRxiv* **2021**, 438252. [CrossRef]

237. Rosenfeld, J.A.; Mason, C.E.; Smith, T.M. Limitations of the Human Reference Genome for Personalized Genomics. *PLoS ONE* **2012**, *7*, e40294. [CrossRef] [PubMed]

238. Mantere, T.; Kersten, S.; Hoischen, A. Long-Read Sequencing Emerging in Medical Genetics. *Front. Genet.* **2019**, *10*, 426. [CrossRef] [PubMed]

239. Leger, A.; Leonardi, T. pycoQC, interactive quality control for Oxford Nanopore Sequencing. *J. Open Source Softw.* **2019**, *4*. [CrossRef]

240. De Coster, W.; D'Hert, S.; Schultz, D.T.; Cruts, M.; Van Broeckhoven, C. NanoPack: Visualizing and processing long-read sequencing data. *Bioinformatics* **2018**, *34*, 2666–2669. [CrossRef] [PubMed]

241. Au, K.F. The blooming of long-read sequencing reforms biomedical research. *Genome Biol.* **2022**, *23*, 1–4. [CrossRef]

242. Ccs: CCS: Generate Highly Accurate Single-Molecule Consensus Reads (HiFi Reads); Github. Available online: https://github.com/PacificBiosciences/ccs (accessed on 15 October 2022).

243. Koren, S.; Walenz, B.P.; Berlin, K.; Miller, J.R.; Bergman, N.H.; Phillippy, A.M. Canu: Scalable and accurate long-read assembly via adaptive *k*-mer weighting and repeat separation. *Genome Res.* **2017**, *27*, 722–736. [CrossRef]

244. Salmela, L.; Walve, R.M.; Rivals, E.; Ukkonen, E. Accurate self-correction of errors in long reads using de Bruijn graphs. *Bioinformatics* **2016**, *33*, 799–806. [CrossRef]

245. Snajder, R.; Leger, A.; Stegle, O.; Bonder, M.J. PycoMeth: A Toolbox for Differential Methylation Testing from Nanopore Methylation Calls. *bioRxiv* **2022**, 480699. [CrossRef]

246. Ni, P.; Huang, N.; Zhang, Z.; Wang, D.-P.; Liang, F.; Miao, Y.; Xiao, C.-L.; Luo, F.; Wang, J. DeepSignal: Detecting DNA methylation state from Nanopore sequencing reads using deep-learning. *Bioinformatics* **2019**, *35*, 4586–4595. [CrossRef]

247. 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. *Nat. Methods* **2018**, *15*, 461–468. [CrossRef]

248. Edge, P.; Bansal, V. Longshot enables accurate variant calling in diploid genomes from single-molecule long read sequencing. *Nat. Commun.* **2019**, *10*, 1–10. [CrossRef]

249. Shafin, K.; Pesout, T.; Chang, P.-C.; Nattestad, M.; Kolesnikov, A.; Goel, S.; Baid, G.; Kolmogorov, M.; Eizenga, J.M.; Miga, K.H.; et al. Haplotype-aware variant calling with PEPPER-Margin-DeepVariant enables high accuracy in nanopore long-reads. *Nat. Methods* **2021**, *18*, 1322–1332. [CrossRef]

250. Heller, D.; Vingron, M. SVIM: Structural variant identification using mapped long reads. *Bioinformatics* **2019**, *35*, 2907–2915. [CrossRef]

251. Heller, D.; Vingron, M. SVIM-asm: Structural variant detection from haploid and diploid genome assemblies. *Bioinformatics* **2020**, *36*, 5519–5521. [CrossRef] [PubMed]

252. Jiang, T.; Liu, Y.; Jiang, Y.; Li, J.; Gao, Y.; Cui, Z.; Liu, Y.; Liu, B.; Wang, Y. Long-read-based human genomic structural variation detection with cuteSV. *Genome Biol.* **2020**, *21*, 1–24. [CrossRef]

253. Freire, B.; Ladra, S.; Parama, J.R. Memory-Efficient Assembly using Flye. *IEEE/ACM Trans. Comput. Biol. Bioinform.* 2021; *online ahead of print*. [CrossRef]

254. Shafin, K.; Pesout, T.; Lorig-Roach, R.; Haukness, M.; Olsen, H.E.; Bosworth, C.; Armstrong, J.; Tigyi, K.; Maurer, N.; Koren, S.; et al. Nanopore sequencing and the Shasta toolkit enable efficient de novo assembly of eleven human genomes. *Nat. Biotechnol.* **2020**, *38*, 1044–1053. [CrossRef] [PubMed]

255. Cheng, H.; Concepcion, G.T.; Feng, X.; Zhang, H.; Li, H. Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. *Nat. Methods* **2021**, *18*, 170–175. [CrossRef] [PubMed]

256. Chin, C.-S.; Peluso, P.; Sedlazeck, F.J.; Nattestad, M.; Concepcion, G.T.; Clum, A.; Dunn, C.; O'Malley, R.; Figueroa-Balderas, R.; Morales-Cruz, A.; et al. Phased diploid genome assembly with single-molecule real-time sequencing. *Nat. Methods* **2016**, *13*, 1050–1054. [CrossRef]

257. Zimin, A.V.; Marçais, G.; Puiu, D.; Roberts, M.; Salzberg, S.L.; Yorke, J.A. The MaSuRCA genome assembler. *Bioinformatics* **2013**, *29*, 2669–2677. [CrossRef]

258. Di Genova, A.; Buena-Atienza, E.; Ossowski, S.; Sagot, M.-F. Efficient hybrid de novo assembly of human genomes with WENGAN. *Nat. Biotechnol.* **2020**, *39*, 422–430. [CrossRef]

259. Vaser, R.; Sović, I.; Nagarajan, N.; Šikić, M. Fast and accurate de novo genome assembly from long uncorrected reads. *Genome Res.* **2017**, *27*, 737–746. [CrossRef]

260. Walker, B.J.; Abeel, T.; Shea, T.; Priest, M.; Abouelliel, A.; Sakthikumar, S.; Cuomo, C.A.; Zeng, Q.; Wortman, J.; Young, S.K.; et al. Pilon: An Integrated Tool for Comprehensive Microbial Variant Detection and Genome Assembly Improvement. *PLoS ONE* **2014**, *9*, e112963. [CrossRef]

261. Chin, C.-S.; Alexander, D.H.; Marks, P.; Klammer, A.A.; Drake, J.; Heiner, C.; Clum, A.; Copeland, A.; Huddleston, J.; Eichler, E.E.; et al. Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. *Nat. Methods* **2013**, *10*, 563–569. [CrossRef]

262. Belser, C.; Istace, B.; Denis, E.; Dubarry, M.; Baurens, F.-C.; Falentin, C.; Genete, M.; Berrabah, W.; Chèvre, A.-M.; Delourme, R.; et al. Chromosome-scale assemblies of plant genomes using nanopore long reads and optical maps. *Nat. Plants* **2018**, *4*, 879–887. [CrossRef]

263. Ballouz, S.; Dobin, A.; Gillis, J.A. Is it time to change the reference genome? *Genome Biol.* **2019**, *20*, 1–9. [CrossRef] [PubMed]

264. Valiente-Mullor, C.; Beamud, B.; Ansari, I.; Francés-Cuesta, C.; García-González, N.; Mejía, L.; Ruiz-Hueso, P.; González-Candelas, F. One is not enough: On the effects of reference genome for the mapping and subsequent analyses of short-reads. *PLoS Comput. Biol.* **2021**, *17*, e1008678. [CrossRef] [PubMed]

265. Kim, H.-S.; Jeon, S.; Kim, C.; Kim, Y.K.; Cho, Y.S.; Kim, J.; Blazyte, A.; Manica, A.; Lee, S.; Bhak, J. Chromosome-scale assembly comparison of the Korean Reference Genome KOREF from PromethION and PacBio with Hi-C mapping information. *GigaScience* **2019**, *8*, giz125. [CrossRef] [PubMed]

266. Ouzhuluobu; He, Y.; Lou, H.; Cui, C.; Deng, L.; Gao, Y.; Zheng, W.; Guo, Y.; Wang, X.; Ning, Z.; et al. De novo assembly of a Tibetan genome and identification of novel structural variants associated with high-altitude adaptation. *Natl. Sci. Rev.* **2019**, *7*, 391–402. [CrossRef]

267. Li, Z.; Chen, Y.; Mu, D.; Yuan, J.; Shi, Y.; Zhang, H.; Gan, J.; Li, N.; Hu, X.; Liu, B.; et al. Comparison of the two major classes of assembly algorithms: Overlap-layout-consensus and de-bruijn-graph. *Brief. Funct. Genom.* **2011**, *11*, 25–37. [CrossRef]

268. Khan, A.R.; Pervez, M.T.; Babar, M.E.; Naveed, N.; Shoaib, M. A Comprehensive Study of De Novo Genome Assemblers: Current Challenges and Future Prospective. *Evol. Bioinform.* **2018**, *14*, 1176934318758650. [CrossRef] [PubMed]

269. Chen, Z.; Erickson, D.L.; Meng, J. Benchmarking hybrid assembly approaches for genomic analyses of bacterial pathogens using Illumina and Oxford Nanopore sequencing. *BMC Genom.* **2020**, *21*, 1–21. [CrossRef] [PubMed]

270. Dierckxsens, N.; Li, T.; Vermeesch, J.R.; Xie, Z. A benchmark of structural variation detection by long reads through a realistic simulated model. *Genome Biol.* **2021**, *22*, 1–16. [CrossRef] [PubMed]

271. Abel, H.J.; Genomics, N.C.F.C.D.; Larson, D.E.; Regier, A.A.; Chiang, C.; Das, I.; Kanchi, K.L.; Layer, R.M.; Neale, B.M.; Salerno, W.J.; et al. Mapping and characterization of structural variation in 17,795 human genomes. *Nature* **2020**, *583*, 83–89. [CrossRef] [PubMed]
272. Lin, J.; Jia, P.; Wang, S.; Ye, K. Comparison and Benchmark of Long-Read Based Structural Variant Detection Strategies. *bioRxiv* **2022**, 503274. [CrossRef]
273. Sakamoto, Y.; Xu, L.; Seki, M.; Yokoyama, T.T.; Kasahara, M.; Kashima, Y.; Ohashi, A.; Shimada, Y.; Motoi, N.; Tsuchihara, K.; et al. Long-read sequencing for non-small-cell lung cancer genomes. *Genome Res.* **2020**, *30*, 1243–1257. [CrossRef]
274. Shiraishi, Y.; Koya, J.; Chiba, K.; Saito, Y.; Okada, A.; Kataoka, K. Precise Characterization of Somatic Structural Variations and Mobile Element Insertions from Paired Long-Read Sequencing Data with Nanomonsv. *bioRxiv* **2021**, 214262. [CrossRef]
275. Berger, M.F.; Mardis, E.R. The emerging clinical relevance of genomics in cancer medicine. *Nat. Rev. Clin. Oncol.* **2018**, *15*, 353–365. [CrossRef]
276. Mardis, E.R. The Impact of Next-Generation Sequencing on Cancer Genomics: From Discovery to Clinic. *Cold Spring Harb. Perspect. Med.* **2018**, *9*, a036269. [CrossRef]
277. Cai, Z.; Poulos, R.C.; Liu, J.; Zhong, Q. Machine learning for multi-omics data integration in cancer. *iScience* **2022**, *25*, 7. [CrossRef]
278. Wang, L.-B.; Karpova, A.; Gritsenko, M.A.; Kyle, J.E.; Cao, S.; Li, Y.; Rykunov, D.; Colaprico, A.; Rothstein, J.H.; Hong, R.; et al. Proteogenomic and metabolomic characterization of human glioblastoma. *Cancer Cell* **2021**, *39*, 509–528.e20. [CrossRef] [PubMed]
279. Fujimoto, A.; Wong, J.H.; Yoshii, Y.; Akiyama, S.; Tanaka, A.; Yagi, H.; Shigemizu, D.; Nakagawa, H.; Mizokami, M.; Shimada, M. Whole-genome sequencing with long reads reveals complex structure and origin of structural variation in human genetic variations and somatic mutations in cancer. *Genome Med.* **2021**, *13*, 1–15. [CrossRef] [PubMed]