

We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

6,100

Open access books available

149,000

International authors and editors

185M

Downloads

Our authors are among the

154

Countries delivered to

TOP 1%

most cited scientists

12.2%

Contributors from top 500 universities



WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com



Chapter

Applications of CRISPR/Cas9 for Selective Sequencing and Clinical Diagnostics

Maximilian Evers, Björn Brändl, Franz-Josef Müller, Sönke Friedrichsen and Stephan Kolkenbrock

Abstract

In this chapter, we will discuss the applications of CRISPR/Cas9 in the context of clinical diagnostics. We will provide an overview of existing methods and their use cases in the diagnostic field. Special attention will be given to selective sequencing approaches using third-generation sequencing and PAM-site requirements. As target sequences in an AT-rich environment cannot easily be accessed by the commercially available SpCas9 due to rarity of NGG PAM-sites, new enzymes such as ScCas9 with PAM-site requirements of NNG will be highlighted. Original research on CRISPR/Cas9 systems to determine molecular glioma markers by enriching regions of interest will be discussed in the context of potential future applications in clinical diagnostics.

Keywords: CRISPR/Cas9, clinical diagnostics, selective sequencing, PAM site, cancer

1. Introduction

1.1 Current diagnostic context

Emerging infectious diseases such as COVID-19, acquired or hereditary genetic defects causing cancer and other illnesses fuel the need for fast and cost-effective diagnostic tools. Gold standard for many types of disease detection is the real-time polymerase chain reaction (PCR) due to its robustness and sensitivity toward molecular biomarkers associated with diseases. Especially, quantitative PCR (qPCR) and reverse transcriptase qPCR have been staples of infectious disease diagnostics to determine the presence of pathogens and viral loads [1, 2], but have also proven their efficacy in tumor diagnostics due to high sensitivity and low input requirements [3, 4]. Many next-generation sequencing (NGS) approaches are also used as diagnostic tools. Whole exome and genome sequencing is used for the investigation of many molecular markers. The benefits of NGS include the ability to screen a large amount of possible target genes in tandem for comparatively low cost [5]. Furthermore, NGS can be used for unbiased detection and species level determination of pathogens in

septic patients. This removes the need for time-consuming blood cultures [6]. Other well-established methods in diagnostics are based on antigen-antibody interaction, such as the Enzyme-Linked Immunosorbent Assay (ELISA) or paper-based lateral flow assays [7, 8].

A new addition to the diagnostic toolbox is clustered regularly interspaced short palindromic repeats (CRISPR)-based diagnostics. CRISPR-associated (Cas) proteins are RNA-guided endonucleases originally part of the adaptive immune system of prokaryotes to ward off invading nucleic acids. Several types of CRISPR/Cas systems have been discovered, and some have been used for diagnostic applications such as Cas12 and Cas13 for methods such as DNA endonuclease-targeted CRISPR *trans* reporter (DETECTR) and specific high-sensitivity enzymatic reporter unlocking (SHERLOCK) and SHERLOCKv2 [9], which were recently developed as potent tools for COVID-19 detection. This chapter will focus on the utility of CRISPR/Cas9 in clinical diagnostics.

1.2 CRISPR/Cas9

CRISPR RNAs (crRNA) provide the targeting mechanism for the Cas9 nuclease activity. crRNAs are hybridized with trans-activating crRNA (tracrRNA), providing a stem-loop structure that anchors the RNA-complex to the Cas9 protein. The crRNA can be engineered to target a wide array of sequences rendering CRISPR/Cas9 a powerful tool for targeted gene editing and recognition. Cas9 proteins are characterized by two nuclease domains forming the active center, HNH and RuvC [10]. HNH is a single nuclease domain responsible for cleaving the DNA strand complementary to the RNA guide. RuvC is split into three subdomains, with RuvC I at the N-terminus of the protein and RuvC II/III flanking the HNH domain near the center of the amino acid sequence [11]. The catalytic residues D10 (in RuvC I) and H840 (in HNH) can be substituted to either limit nuclease activity in case of a single-site inactivation to create a Cas9 nickase or to generate a catalytically inactive/dead Cas9 (dCas9) variant in case of a double-site inactivation [12]. In addition to the nuclease domains, Cas9 possesses a protospacer adjacent motif (PAM)-interacting (PI) domain. The PAM is a short nucleic acid sequence downstream of the crRNA conferred target sequence, required for nuclease activity and target sequence interrogation [13]. It is thought to have originated in prokaryotes so as not to target their own DNA and thus to prevent an autoimmune response [14]. While the sequence to be cut can be easily defined via crRNA, the obligatory requirement of a protospacer adjacent motif (PAM) sequence next to the target sequence [15, 16] limits the applications of Cas9 in clinical diagnostics. Due to this limitation, regions of interest without matching PAM-site cannot be cleaved and subsequently analyzed. Several variants of Cas9 enzymes have been generated to partially circumvent those limitations with a relaxation of the PAM-site requirement. The Cas9 from *Streptococcus pyogenes* (SpCas9) natively recognizes the PAM 5'-NGG-3' but was modified (termed xCas9) to accept a broad range of PAM sites, including 5'-NG-3', 5'-GAA-3', and 5'-GAT-3' [17]. Additionally, Cas9 enzymes from different hosts such as the Cas9 from *Streptococcus canis* (ScCas9) were modified to be more promiscuous regarding PAM site recognition (termed Cas9-SC++), now accepting 5'-NNG-3' as a PAM site [18]. A Cas9 homolog discovered in *Francisella novicida* (FnCas9) also recognizes the 5'-NGG-3' PAM but was successfully engineered to accept a 5'-YG-3' PAM [19].

1.3 Cas9 in diagnostic methods

In the area of molecular diagnostics, CRISPR/Cas9 systems have proven to be effective tools in distinguishing between different Zika virus strains. Pardee et al. (2016) [20] used nucleic acid sequence-based amplification (NASBA) in combination with Zika strain-specific sgRNA/Cas9 and toehold switches to create a colorimetric assay to detect and differentiate African and American Zika virus strains. A toehold switch is an RNA molecule combining a sensor and a reporter sequence. Without the presence of the trigger, an RNA molecule complementary to the sensor sequence, a hairpin structure is formed. It limits access to the ribosomal binding site and therefore inhibits translation of the reporter. Due to strand displacement upon hybridization with the trigger RNA, the hairpin structure is resolved, allowing the translation of the reporter [21]. The toehold switch was designed to regulate *lacZ* expression and was activated by the Zika virus RNA amplicons, which allowed for colorimetric *in vitro* detection of the target RNA. Due to sequence differences, PAM site locations vary between the strains, which was exploited for targeted truncation of RNA amplicons of only one strain in a method termed NASBA-CRISPR Cleavage (NASBACC). Truncated RNA amplicons could not activate the toehold switch, which allowed for discrimination between the strains [20].

CRISPR-Cas9 nickase (SpCas9H840A nickase) strand displacement amplification (CRISDA) is an ultrasensitive method to detect target DNA with single-nucleotide accuracy and attomolar sensitivity. A pair of SpCas9 nickase ribonucleoproteins (RNPs) introduce nicks in the flanking areas next to the region of interest. Initial primers anneal to the nicked strands from where strand displacement amplification begins. Biotin and Cy5-labeled peptide nucleic acid (PNAs) probes are introduced to the amplification mix to detect and quantify amplicons. The PNA binds to the amplicons and enables a pulldown using streptavidin-coated magnetic beads. Fluorescence measurement of the pulled-down DNA allows for quantification of the generated amplicons [22].

Another nucleic acid detection strategy is CRISPR/Cas9-triggered isothermal exponential amplification reaction (CAS-EXPAR). It is based on CRISPR/Cas9 cleavage and nicking endonuclease (NEase)-mediated nucleic acids amplification. Cas9 cleavage of the target produces a primer for the CAS-EXPAR reaction, wherein the target “X” hybridizes with a construct containing two sequences complementary to the target (“X”), which are connected via a PAMmer. Upon extension of the double strand, Cas9 cleaves off the newly synthesized DNA, which in turn acts as a primer itself. This strategy was shown to have a detection limit of 0.82 amol and high specificity, discriminating single-base mismatches [23].

Lateral flow assays are state of the art in point-of-care diagnostics. CRISPR/Cas9-mediated lateral flow nucleic acid assay (CASLFA) combines the sensitivity of Cas9 endonuclease with the ease of use of lateral flow assays. CASLFA was developed for the identification of *Listeria monocytogenes*, different genetically modified organisms (GMOs), and the African swine fever virus (ASFV) [24].

FnCas9 editor-linked uniform detection assay (FELUDA) is a diagnostic tool combining preamplification of a target sequence using biotinylated primers with inactive FnCas9 to detect target sequences. The used tracrRNA is FAM-labeled and can be recognized via antibodies, and the capture of target sequences is paired with a lateral flow readout. The biotinylated amplicons bind to the test region via

streptavidin interaction. If FnCas9 binds to the amplicon DNA, it will be retained in the test region, allowing for antibody-based detection in the form of a visible band. dFnCas9 was used for this assay, as it exhibits lower affinity toward sequences with single-nucleotide mismatches to the crRNA used than SpCas9 [25].

Finding Low Abundance Sequences by Hybridization (FLASH) is a method that combines Cas9 digestion, PCR, and Illumina sequencing to detect and identify antimicrobial resistance in microbial DNA samples. Isolated DNA is dephosphorylated before Cas9 digestion of target sequences. The double-strand breaks introduced by Cas9 remain phosphorylated and are subsequently dA-tailed. Adapters are ligated to the dA-tailed target sequences, which are then amplified via PCR. The resulting library can be sequenced via Illumina sequencing and achieve sub-attomolar sensitivity [26].

Next to infectious disease detection, another field of interest for targeted Cas9 diagnostics is cancer, one of the world's leading causes of premature death [27]. As cancerous unregulated cell growth can be caused by a combination of genetic defects, it is vital for prognosis and treatment to accurately diagnose its molecular cause. Cancer diagnostics currently is often based on histological analysis of tumor tissue. Because histology is predetermined by genetics, research efforts to quickly identify aberrant tumor marker genes on the molecular level are being pursued.

One application to potentially target this challenge is CRISPR-Chip. This method utilizes dCas9 immobilized on a graphene surface, acting as a conductor. The dCas9 is paired with a specifically designed sgRNA to recognize its target. Upon binding target DNA, the conductivity of the immobilized dCas9 changes, which can be measured via the graphene surface. This allows for detection limits of 1.7 fM gDNA. Though it was demonstrated with target sequences associated with Duchenne muscular dystrophy, it could be used for any sequences as long as a suitable PAM-site is flanking the region of interest [28].

Another route to follow in molecular tumor diagnosis is the sequencing of tumor marker genes. With the advent of second- and third-generation sequencing, the feasibility of sequencing approaches in standard diagnostics is increasing due to lower costs and shorter sequencing times. However, a combination with CRISPR/Cas technology allows for a specific sequencing of the regions of interest, boosting the output of relevant regions, and thus enabling a faster and very specific and sensitive sequencing approach.

1.4 Tumor biomarker selection for Cas9-targeted sequencing

To maximize utility of Cas9-targeted sequencing, biomarkers such as mutations or methylation patterns with defined locations are favorable. Because sequencing times are determined by target sequence length, biomarkers such as defined SNPs allow for higher throughput, as flanks of the targeted sequence can be chosen in close proximity to the region of interest. In our research we developed an amplification-independent workflow to assess the tumor marker status of six relevant genes/regions in brain tumors following the 2021 WHO Classification of Tumors of the Central Nervous System [29]. These genes/regions, their function, and glioma-relevant mutations are described in the following.

Isocitrate dehydrogenases 1 and 2 (IDH1, IDH2) are crucial enzymes that catalyze the oxidative decarboxylation of isocitrate to α -ketoglutarate during the Krebs cycle. Common mutations associated with glioma formation are related to codons 132 for IDH1 and 172 for IDH2 causing aberrant enzymatic activity and in turn the accumulation of 2-hydroxyglutarate, which inhibits many α -ketoglutarate dependent enzymes such as DNA-demethylases, leading to DNA hypermethylation [30].

Additionally, the promoter of telomerase reverse transcriptase (pTERT) represents a clinically relevant target due to its close association with oncogenesis and immortalization of cell lines [31]. The mutations C228T and C250T are commonly associated with aberrant expression patterns as these mutations create *de novo* binding sites for members of the E26 transformation-specific family of transcription factors [31].

H3F3A and *Hist1H3B* encode histone subunits H3.3 and H3.1, respectively. K27M variants are observed in different cancer types, such as Diffuse Intrinsic Pontine Glioma, and G34R/V substitution in H3.3 is also associated with young adult high-grade astrocytoma [32, 33].

BRAF encodes a member of the Raf kinase family, B-Raf, and is a growth signal transduction protein kinase that regulates pathways associated with cell division and differentiation. The amino acid substitution V600E of B-Raf increases its basal activity and stimulates cell division and differentiation pathways. This is associated with a variety of different cancer types [34].

As mutation detection via sequencing is of interest, it is crucial to be aware of the benefits and drawbacks of the used sequencing technologies for the development of a medical diagnostic application.

1.5 Current sequencing technologies

Currently, the most widely used next-generation sequencing technologies on the market are Illumina short-read sequencing, PacBio (also referred to as Single-Molecule Real-Time (SMRT) sequencing), 454 pyrosequencing, ion-torrent/proton sequencing, and nanopore sequencing by Oxford Nanopore Technologies (ONT) [35]. Illumina, 454, and ion-torrent sequencing technologies are referred to as second-generation sequencing technologies. They deliver short reads of about 50–1000 bp in length and their parallelization in sequencing reaction results in a high read throughput (0.7–15 M reads per run) and an amount of sequence information of about 0.5–8.5 Gb per run [36]. PacBio and nanopore sequencing are referred to as third-generation sequencing technologies. They usually deliver tens of kb per read up to several Mb, but far fewer reads in total. For nanopore sequencing, the amount of sequence information is strongly dependent on the flow cell. Depending on the nucleic acid library preparation and its quality, up to 2.8 Gb per run is theoretically achievable on a Flongle flow cell, 10–15 Gb on a MinION flow cell [37], and up to 153 Gb per run was reported by using the PromethION flow cell [38]. PacBio sequencing utilizes SMRT cells for sequencing, usually generating 55,000–365,000 reads per run with an average read length of 10–16 kb [39] and 15–96 Gb per run [40].

Nanopore and PacBio sequencing allow for real-time sequencing with parallel base calling of the steadily increasing raw sequencing information allowing direct usage of the results during the run. In addition, both techniques allow detection of epigenetic information of each nucleotide sequenced [41], which can be a piece of important additional information in clinical cancer diagnostics and treatment [42–44]. While methylation of a base directly impacts the raw signal of the nanopore sequencing and thus can be distinguished from an unmodified nucleotide, PacBio detects methylation by a change in DNA-polymerase kinetics during synthesis. Due to the “sequencing by synthesis” technology of second-generation sequencing techniques, they cannot detect epigenetic modification directly, but only via a pretreatment step such as bisulfite treatment, endonuclease digestion, or affinity enrichment [45].

In summary, the selection of the sequencing technology used for clinical diagnostics will be strongly dependent on requirements such as mode of analysis and time to

Sequencing technology	Real-time	Average sequence length	Typical number of reads	Amount sequence data per run	Methylation status detection
Illumina	no	2 × 150 bp (HiSeq 4000) [39]	5 billion (HiSeq 4000) [39]	1300–1500 Gb*	no
		2 × 300 bp (MiSeq) [39]	25 million (MiSeq) [39]	4.5–5.1 Gb*	
454 pyro-sequencing	no	400–1000 bp [46]	> 1 million [32]	35–450 Mb [46]	no
Ion-torrent	no	200–600 bp*	2–130 million*	0.3–50 Gb*	no
PacBio	yes	10–16 kb [39]	55,000–365,000 [25]	15 Gb (Sequel) [40]	yes
				96 Gb (Sequel II) [40]	
Oxford Nanopore	yes	10–30 kb (all flow cells) record 2.6 Mb [47]	200,000 (Flongle) [48]	Up to 2.8 Gb (Flongle)*	yes
			1,200,000 (MinION) [48]	10–15 Gb (MinION) [37]	
			Up to 250 million (PromethION)*	Up to 153 Gb (PromethION) [38]	

*Specifications obtained from the respective company's website (as of 01.06.2022).

Table 1.
Most widely used sequencing technologies and their characteristics.

results. The characteristics in this respect of each mentioned sequencing technology are summarized in **Table 1**.

Next to Illumina sequencing-based methods such as FLASH, third-generation sequencing can also be paired with Cas9 enrichment of target sequences. PacBio uses a generic SMRT sequencing library, which is digested by Cas9. The digested sequences are then ligated to a second set of adapters, which is used for magnetic-bead-based separation of the targeted sequences, allowing for target sequence enrichment [49].

For ONT's Cas9-targeted sequencing process (nCATS), DNA is dephosphorylated before Cas9 digestion. Like FLASH, the phosphorylated ends of the cleaved DNA are dA-tailed and ligated to sequencing adapters, allowing for selective sequencing [50]. A similar approach was pursued in the following experimental section to develop a CRISPR/Cas, third-generation sequencing assay for diagnosis of. Utilizing the promising properties of nanopore and Cas9-dependent target enrichment, we developed an amplification-independent workflow to assess glioma biomarkers.

2. Development of a CRISPR/Cas9-targeted sequencing approach

2.1 Material and methods

To test the feasibility of nanopore sequencing in brain tumor marker detection, we used pUC57 vectors containing 2 kb target sequence of a given tumor marker as

either wild-type or containing a clinically relevant mutation. Cas9-RNP populations were prepared to cleave the DNA upstream and downstream of a given mutation site. The excised double-stranded DNA was used for sequencing library preparation using the SQK-CS9109 Cas9 sequencing Kit from ONT. Flongle flow cells (version R.9.4.1) were used for sequencing. Sequences were assessed using a minimap2 [51] alignment followed by custom SNP calling using python scripts. As tumor treatment is very time-sensitive [52], the possibility of intra-surgical diagnostics could alleviate an unmet clinical need. Therefore, we evaluated the results not only by accuracy but also regarding time to results. In addition to the complete sequencing data acquired a subset generated during the first 15 min of sequencing was also used for analysis.

2.2 Cas9-RNP preparation

Alt-R® S.p. Cas9 Nuclease V3 including tracrRNA and crRNAs were purchased from Integrated DNA Technologies (Coralville, IA, USA). The crRNAs were designed to target at least 200 bp upstream and downstream of each mutation site resulting in at least 1000 bp of excised dsDNA in total. crRNAs were designed for *IDH1*, *IDH2*, pTERT, *H3F3A*, *Hist1H3B*, and *BRAF*. All sequences of used crRNAs are given in **Table 2**. To anneal crRNA and tracrRNA, 8 µL Duplex Buffer (IDT), 1 µL tracrRNA (100 µM), and 1 µL crRNA Pool (100 µM, equimolar) were assembled in 0.2 mL thin-walled PCR tubes and incubated at 95 °C in a thermal cycler. The mix was allowed to cool to room temperature (RT) afterward. The annealed crRNA/tracrRNA (10 µM) was added to 79.2 µL nuclease-free water, 10 µL Reaction buffer (SQK-CS9109 Kit), and 0.8 µL Alt-R® S.p. Cas9 Nuclease V3 (62 µM) and mixed thoroughly by flicking. The RNPs were formed by incubation at RT for 30 min and stored at 4°C until needed. Two different RNP populations were prepared with different crRNAs. Population 1 included all crRNAs described in **Table 2**, whereas population 2 was prepared with only the two crRNAs targeting *IDH1*.

2.3 Cas9 digestion of pUC57 plasmids

The pUC57 vectors containing tumor marker sequences used as target DNA for Cas9 digestion were purchased from GenScript (Piscataway, NJ, USA). For DNA digestion and library preparation, the SQK-CS9109 Cas9 sequencing kit from ONT (Oxford, UK). DNA digestion was set up by adding template plasmids to Cas9 RNPs, reaction buffer, dATP, and Taq DNA-polymerase. One digestion was performed with

Target	crRNA Seq fw (5'-3')	crRNA Seq rv (5'-3')	Fragment size [bp]
IDH1	ATGTTTAATACAATCTTTGG	GCTTCCCATTACAAGAGGAG	1062
IDH2	AGTGCACACGATGTTTCTGC	TCGTCCTCACGACAACACTT	1601
pTERT	CTCCCTGACGCTATGGTTCC	GTCAAGGAGCCCAAGTCGCG	1443
H3F3A	AATTTGACTCGACCTTCCAG	TATTTGCGGAGGCTAAGTCT	1081
Hist1H3B	GCATTCCTAACTATCTTGAA	CATAGTCTAATGCTTTCCGG	1334
BRAF	GACCCTCTAAAACGGGTGTGA	GCATGCATGTATAGGAGAGC	1584

Table 2.

List of crRNAs used for enzymatic excision of tumor marker DNA from pUC57 vectors containing target gene fragments.

	Digestion 1	Digestion 2
pUC57::IDH1_Wt (80 ng/ μ L)	—	2 μ L
pUC57::IDH1_R132C (80 ng/ μ L)	1 μ L	—
pUC57::IDH1_R132H (80 ng/ μ L)	—	2 μ L
pUC57::IDH2_Wt (80 ng/ μ L)	1 μ L	—
pUC57::pTERT_Wt (80 ng/ μ L)	1 μ L	—
pUC57::H3F3A_Wt (80 ng/ μ L)	1 μ L	—
pUC57::Hist1H3B_Wt (80 ng/ μ L)	1 μ L	—
pUC57::BRAF_Wt (80 ng/ μ L)	1 μ L	—
ddH ₂ O	21 μ L	23 μ L
Cas9 RNPs (Population 1) (10 μ M)	10 μ L	—
Cas9 RNPs (Population 2) (10 μ M)	—	10 μ L
Reaction Buffer	3 μ L	3 μ L
dATP	1 μ L	1 μ L
Taq Polymerase	1 μ L	1 μ L

Table 3.
Setup for digestion of different pUC57 plasmid mixtures using different Cas9-RNP populations.

a mixture of six plasmids (80 ng each), each containing a different marker, in order to test multiplexing. Another digestion was set up with a mixture of two plasmids (160 ng each) containing different mutations of the same marker, *IDH1*, in order to test variant calling capabilities. The different reaction mixtures were prepared as shown in **Table 3**.

The reaction mixtures were incubated for 30 min at 37°C for Cas9 cleavage. Subsequently it was incubated at 72°C for 5 min for Taq Polymerase facilitated dA-tailing of cleaved fragments.

2.4 Library preparation

For sequencing adapter ligation to the dA-tailed fragments, the digested and dA-tailed DNA was added to a mixture of 20 μ L Ligation Buffer, 3 μ L nuclease-free water, 10 μ L T4 Ligase, and 5 μ L Adapter Mix. The ligation components were mixed by flicking, spun down, and incubated at RT for 10 min. DNA was purified using AMPure XP beads (Beckman Coulter, Brea, CA, USA). In total, 48 μ L of magnetic beads was added to each sample and mixed by inversion. The samples were incubated for 10 min at RT without agitation. Afterward, samples were spun down, and beads were separated magnetically. The supernatant was discarded, and the pellet washed twice with 250 μ L Short Fragment Buffer (SFB). The wash step consisted of resuspension in SFB and subsequent magnetic separation of the washed beads. After the supernatant was removed, the pellet was resuspended in 13 μ L 50 °C Elution Buffer. The elution mixture was incubated at 50 °C and 1000 rpm in a heater shaker for 10 min. After magnetic separation, 13 μ L of the eluate was removed, and DNA content and purity were analyzed via Nanodrop. Sequencing libraries were created by combining 37.5 μ L Sequencing Buffer, 25.5 μ L Loading Beads, and 12 μ L DNA Library.

2.5 Nanopore sequencing

In total, 37.5 μ L of each library was used for sequencing on a Flongle flow Cell (R9.4.1). DNA contents were 48.6 ng for sample 1, containing fragments of all six plasmids, and 17.4 ng for sample 2, containing fragments of two plasmids, resembling two *IDH1* variants. Sequencing was concluded after 18 h, and a subset of sequences generated during the first 15 min was separated.

2.6 In silico analysis

To analyze the possible mutation sites bioinformatically, alignment references of the tumor marker sequences were created. In the reference sequences, the possible mutation site was deleted; therefore, alignment of each generated sequence produced an insertion mutation during variant calling. The bases recognized as insertions were used to distinguish wild-type and mutated sequences. Each generated sequence was aligned with all possible target references using minimap2 [51]. Subsequently, paftools.js was used for variant-calling [51] and custom scripts accumulated the numbers of wild-type and mutant reads in real time. To ensure the highest possible accuracy, matches between generated sequences and references were split into *mapped generated sequences* and *sequences with tumor marker information*. Because truncated sequences or erroneously sequenced DNA molecules can be aligned to a given reference but yield no tumor marker information, only generated sequences that can unambiguously be identified as a tumor marker variant were used for analysis.

2.7 Results and discussion

Creating a sequencing library from a mixture of six plasmids containing different tumor marker genes enabled us to identify each target with high accuracy, as seen in **Table 4**. These results were achieved after 18 h of sequencing with a library containing

Target	Inspected variation	Mapped generated Sequences	Sequences with tumor marker information	Correctly Identified Sequence variants [%]
IDH1_R132C	R132	128416	74002	99.57
IDH2_Wt	R172	196040	151329	97.59
pTERT_Wt	C228	48932	32613	99.45
	C250		34689	99.61
H3F3A_Wt	K27	39232	21899	99.11
	G34		22308	97.50
Hist1H3B_Wt	K27	45560	19458	99.66
BRAF_Wt	V600	52387	35354	98.36

Table 4.

A sequencing library was prepared from tumor marker DNA excised from synthetic plasmids. Equal amounts of plasmid were used for each target. **Mapped sequences** were identified as a given marker sequence via minimap2, but only **sequences with tumor marker information** were able to be used for SNP calling. Shown is the cumulative output after 18 h of sequencing on a Flongle flow cell.

Target	Inspected variation	Mapped generated Sequences	Sequences with tumor marker information	Correctly Identified Sequence variants [%]
IDH1_R132C	R132	914	604	99.50
IDH2_Wt	R172	894	757	98.94
pTERT_Wt	C228	348	257	100
	C250		269	100
H3F3A_Wt	K27	239	157	100
	G34		166	96.99
Hist1H3B_Wt	K27	292	143	100
BRAF_Wt	V600	246	185	100

Table 5.

A sequencing library was prepared from tumor marker DNA excised from synthetic plasmids. Equal amounts of plasmid were used for each target. **Mapped sequences** were identified as a given marker sequence via *minimap2*, but only **sequences with tumor marker information** were able to be used for SNP calling. Shown is the cumulative output after 15 min of sequencing on a Flongle flow cell.

48.6 ng of plasmid DNA. Overall yield of sequences was high with >39000 generated reads for all markers. *IDH1* and *IDH2* were outliers in this regard as they yielded 128416 and 196040 reads, respectively. As expected, not all generated sequences carried tumor marker information. The ratio of sequences with tumor marker information to all mapped sequences was 42–77%, depending on the given marker (**Table 4**). The ratio of correctly annotated tumor marker variants was >97% in all cases. We demonstrated the specificity of Cas9 cleavage by including a plasmid without target sequences as background in separate control experiments. No sequences derived from this control plasmid were generated during subsequent ONT sequencing.

Accuracy for simulated homozygosity and coverage of each marker after 18 h was very high. As time-to-result is a central parameter in clinical diagnostics, these results were examined regarding the coverage and accuracy after 15 min of sequencing time. This subset revealed a coverage for each marker of >200x after 15 min. Ratios of sequences with tumor marker information out of all generated sequences were similar after 15 min as compared with 18 h. They ranged from 48 to 84%, depending on the observed marker. Accuracy regarding the identification of the tumor marker sequence was also comparably high in this data subset. Between 96.99 and 100% of sequences with tumor marker information were annotated correctly, as shown in **Table 5**.

As these results were achieved with one variant per marker, no conclusions regarding heterozygosity detection were possible. Most mutations are heterozygous. Therefore, a subsequent experiment was performed to assess the analysis accuracy when including a simulated heterozygous mutation. For this experiment, a mixture of equal amounts of pUC57::*IDH1*_Wt and pUC57::*IDH1*_R132H was used for Cas9 sequencing. In total, 17.4 ng of plasmid DNA contained in the library was loaded onto the Flongle flow cell. Results shown in **Table 6** were achieved after 18 h of sequencing, with a subset of sequences generated during the first 15 min of sequencing being evaluated separately. In total, 625 reads were generated after 15 min of which 490 (72%) yielded tumor marker information. After 18 h, 29072 reads were generated and 20875 (78%) yielded tumor marker information. It was found that both *IDH1* variants present in the digested plasmid mix were detected during sequencing. The expected

	Sequencing time 15 min	Sequencing time 18 h
Mapped generated sequences	625	29072
Sequences with tumor marker information	490	20875
Expected ratio (Wt/R132H/other IDH1) [%]	50/50/0	50/50/0
Achieved ratio (Wt/R132H/other IDH1) [%]	40.2/59.59/0.21	38.49/61.4/0.01

Table 6.

A sequencing library was prepared from tumor marker DNA excised from synthetic plasmids. Equal amounts of plasmid were used for each target. **Mapped sequences** were identified as a given marker sequence via minimap2, but only **sequences with tumor marker information** could be used for SNP calling. Shown is the cumulative output after 15 min and 18 h of sequencing on a Flongle flow cell.

ratio of these variants was 50% IDH1_Wt and 50% IDH1_R132H, but an approximated 40%/60% split was observed after 15 min and 18 h.

Although Cas9 selective sequencing focuses on the sequence of interest and thus may lead to faster and more accurate results as compared with a whole genome sequencing approach, there exist some challenges. In a direct sequencing approach, there is no amplification step involved when analyzing only a single mutated base or area. Thus, one complete genome delivers only one read of the desired area, which in case of carcinoma mutations is mainly haploidic. Therefore, a huge amount of highly pure, high molecular weight, genomic DNA must be prepared and used, which depending on the amount of sample and method of nucleic acid preparation may be a limitation. Typically, the usage of 1–10 µg of human genomic DNA for Cas9 digestion is suggested (nCATS, [50]), corresponding to 150.000–1.500.000 copies of a diploid (female) genome, ideally resulting in the same number of reads. However, one must consider inefficiencies in Cas9 digestion, nucleic acid purification, and library preparation together with possible off-target digestion effects. Further on, there exists an intrinsic error rate of each sequencing method used and in case of cancerous tissue, it may consist of a mixture of wild-type and mutated cells depending on tumor heterogeneity and the general quality of tissue sampling. This may result in a general low number of reads, possibly beyond the coverage needed to safely identify a mutation.

Despite these drawbacks, we believe an nCATS-based approach to intra-surgical determination of a molecular tumor marker panel is justified, as it allows for live detection of marker variants including epigenetic information [50]. Preamplification of the targets might alleviate the high input DNA requirements but removes the ability to determine epigenetic properties of the sequences. That would render the effective analysis of markers such as MGMT methylation status, a predictive biomarker for efficacy of chemotherapy [53], impossible. PCR-based approaches such as qPCR would be very sensitive, as even a few copies of target DNA can produce a positive signal [54], but primer sets that incorporate the putative mutation site would be necessary to distinguish between wild-type and mutant sequences. This is a drawback in comparison to the chosen nCATS approach as this only detects anticipated mutations. Immunodetection of possible auto-antibodies (e.g., with ELISA) has been reported to be prone to false positives [55] and even though nanopore sequencing itself is prone to sequencing errors, they are distributed across the sequence, which leads to high consensus accuracy [56]. Drawbacks are the low resolution of homopolymers, which are prone to sequencing errors with the current flow cell generations. Second-generation sequencing would allow for high sensitivity and accuracy and is well-established but delivers only short sequences. Due to its sequencing by synthesis approach, epigenetic information is lost

Diagnostic tool	Mode of action	Advantages	Challenges	Accuracy	Sensitivity
nCATS [50]	Excision of target sequences from genome and subsequent sequencing	Native sequencing—epigenetic profile accessible, real-time basecalling	Large input DNA requirements, possible off-target sequencing	Medium, off-target digestions possible [58]; Nanopore sequencing, high intrinsic error rate [56]	Medium
(Reverse transcriptase) real-time PCR [3, 4]	Detection by amplification	Quantitative sequencing, well-established method	Reliant on primer specificity, affected by mutations in binding regions of primer probes	Very high	Very high
ELISA [7, 59]	Immuno-detection	Can detect a wide range of antigens and antibodies	Prone to false positives [55]	Low–medium	Medium
2nd Gen. Sequencing [57, 60]	Sequencing by synthesis	Established methods, high read count	Short read sequencing, high number of reads needed for contig generation	Very high	Very high
3rd Gen. Sequencing [37, 61]	ONT: electrical current changes PacBio: Fluorescence detection	Long read sequencing, easy assembly of long contigs	High intrinsic error rate (single sequence), low homopolymer resolution	Medium	Medium

Table 7.
Comparison of current diagnostic tools with nCATS.

in this case as well [57]. A comparison between mode of action, advantages and disadvantages, accuracy and sensitivity of those diagnostic tools is shown in **Table 7** below.

3. Conclusion

Considering the number of reads generated in 15 min using Flongle flow cells, the process might be sped up with the use of MinION flow cells. This would increase cost but cut sequencing time in a trade-off to be considered on a case-by-case basis. We demonstrated that the defined plasmid sequences could be analyzed via the described workflow. As the described experiments represent a work in progress, the use of isolated gDNA with significantly lower target sequence density as a template must be demonstrated. Previous works used either amplification-independent whole-genome sequencing (WGS) or amplicon sequencing to assess brain tumor marker variants. The WGS alternative is significantly slower due to the excess of non-target sequences generated but yielded additional epigenetic information for the regions of interest [62]. Enrichment of target sequences via PCR yields more favorable library compositions but eliminates epigenetic information [62]. As shown here, our CRISPR/

Cas9-based approach to enrich native target sequences might be able to combine the advantages of both previous strategies.

The results of the simulated heterozygosity were ~10% off from the expected 50% distribution of IDH1 Wt/IDH1 R132H, as shown in **Table 6**, but the fact that negligible amounts of other IDH1 mutations and no other tumor marker sequences were found is promising toward applications in clinical environments.

Ultimately, the goal of such a workflow should be a time-to-result below the time required for a neurosurgical tumor resection via craniotomy. This way, neurosurgeons could make informed decisions about the extent of the ongoing surgery and initiate personalized therapeutic modalities based on clinically actionable prognostic biomarkers [63]. Provided the RNPs are prepared in advance, the workflow described using the ONT Cas9 Sequencing kit, including 15 min time required for sequencing, would take 1:45 h altogether. Assuming gDNA extraction and preparative dephosphorylation add another 30–45 min [64], the total time-to-results may be as low as 2.5 h. Considering the lower abundance of target sequences in gDNA compared with synthetic plasmids, sequencing times would likely increase to generate the same coverage, which must be accounted for. But this could partly be mitigated by the usage of a MinION flow cell instead of a Flongle flow cell. The analyzed marker panel can be amended by adding or subtracting crRNAs to target different sequences. A prerequisite for this approach is the presence of PAM sites in the vicinity of the new target genes. Mutations in AT-rich regions might be hard to access via SpCas9 because of its PAM-site requirement of 5'-NGG-3'. For this reason, different Cas9 proteins might be suitable candidates for amended workflows, such as xCas9, ScCas9-SC++, or an engineered FnCas9 [17–19]. In summary, these proof-of-concept results suggest that Cas9-aided targeted sequencing can generate diagnostically relevant tumor marker information in a short period of time and therefore might be a feasible diagnostic method for intra-surgical tumor diagnostics.

Acknowledgements

The authors gratefully thank Prof. Dr. Bruno Moerschbacher (University of Münster) and Prof. Dr. Wolfgang Streit (University of Hamburg) for their in-depth discussions and scientific input. We would also like to thank Sabina Schaal for the project management and scientific input. We gratefully acknowledge the financial support of the Federal Ministry of Education and Research (Funding indicator: 13GW0347A).

IntechOpen

Author details

Maximilian Evers¹, Björn Brändl², Franz-Josef Müller³, Sönke Friedrichsen¹
and Stephan Kolkenbrock^{1*}


1 Altona Diagnostics GmbH, Hamburg, Germany

2 Max-Planck-Institute of Molecular Genetics, Berlin, Germany

3 University Hospital Schleswig Holstein, Kiel, Germany

*Address all correspondence to: stephan.kolkenbrock@altona-diagnostics.com

IntechOpen

© 2022 The Author(s). Licensee IntechOpen. This chapter is distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/3.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. 

References

- [1] Improved RT-qPCR could transform diagnostics [Internet]. 2020. Available from: <https://www.nature.com/articles/d42473-020-00424-1>. [Accessed: 2022-07-07]
- [2] el Jaddaoui I, Allali M, Raoui S, Sehli S, Habib N, Chaouni B, et al. A review on current diagnostic techniques for COVID-19. *Expert Review of Molecular Diagnostics*. 2021;**21**:141-160. DOI: 10.1080/14737159.2021.1886927
- [3] Skrzypski M. Quantitative reverse transcriptase real-time polymerase chain reaction (qRT-PCR) in translational oncology: Lung cancer perspective. *Lung Cancer*. 2008;**59**:147-154. DOI: 10.1016/j.lungcan.2007.11.008
- [4] Ståhlberg A, Zoric N, Åman P, Kubista M. Quantitative real-time PCR for cancer detection: The lymphoma case. *Expert Review of Molecular Diagnostics*. 2005;**5**:221-230. DOI: 10.1586/14737159.5.2.221
- [5] Matthijs G, Souche E, Alders M, Corveleyn A, Eck S, Feenstra I, et al. Guidelines for diagnostic next-generation sequencing. *European Journal of Human Genetics*. 2016;**24**:2-5. DOI: 10.1038/ejhg.2015.226
- [6] Grumaz S, Stevens P, Grumaz C, Decker SO, Weigand MA, Hofer S, et al. Next-generation sequencing diagnostics of bacteremia in septic patients. *Genome Medicine*. 2016;**8**:1-13. DOI: 10.1186/s13073-016-0326-8
- [7] Tighe PJ, Ryder RR, Todd I, Fairclough LC. ELISA in the multiplex era: Potentials and pitfalls. *PROTEOMICS—Clinical Applications*. 2015;**9**:406-422. DOI: 10.1002/prca.201400130
- [8] Koczula KM, Gallotta A. Lateral flow assays. *Essays in Biochemistry*. 2016;**60**:111-120. DOI: 10.1042/EBC20150012
- [9] Mustafa MI, Makhawi AM. Sherlock and detectr: CRISPR-cas systems as potential rapid diagnostic tools for emerging infectious diseases. *Journal of Clinical Microbiology*. 2021;**59**:e00745-e00720. DOI: 10.1128/JCM.00745-20
- [10] Zuo Z, Liu J. Structure and dynamics of Cas9 HNH domain catalytic state. *Scientific Reports*. 2017;**7**:1-13. DOI: 10.1038/s41598-017-17578-6
- [11] Hsu PD, Lander ES, Zhang F. Development and applications of CRISPR-Cas9 for genome engineering. *Cell*. 2014;**157**:1262-1278
- [12] Ma Y, Zhang L, Huang X. Genome modification by CRISPR/Cas9. *The FEBS Journal*. 2014;**281**:5186-5193. DOI: 10.1111/febs.13110
- [13] Hsu PD, Scott DA, Weinstein JA, Ran FA, Konermann S, Agarwala V, et al. DNA targeting specificity of RNA-guided Cas9 nucleases. *Nature Biotechnology*. 2013;**31**:827-832. DOI: 10.1038/nbt.2647
- [14] Mali P, Esvelt KM, Church GM. Cas9 as a versatile tool for engineering biology. *Nature Methods*. 2013;**10**:957-963. DOI: 10.1038/nmeth.2649
- [15] Jiang F, Doudna JA. CRISPR–Cas9 structures and mechanisms. *Annual Review of Biophysics*. 2017;**46**:505-529. DOI: 10.1146/annurev-biophys-062215-010822
- [16] Chen YC, Sheng J, Trang P, Liu F. Potential application of the CRISPR/

CAS9 system against herpesvirus infections. *Viruses*. 2018;**10**:291. DOI: 10.3390/v10060291

[17] Hu JH, Miller SM, Geurts MH, Tang W, Chen L, Sun N, et al. Evolved Cas9 variants with broad PAM compatibility and high DNA specificity. *Nature*. 2018;**556**:57-63. DOI: 10.1038/nature26155

[18] Chatterjee P, Jakimo N, Lee J, Amrani N, Rodríguez T, Koseki SRT, et al. An engineered ScCas9 with broad PAM range and high specificity and activity. *Nature Biotechnology*. 2020;**38**:1154-1158. DOI: 10.1038/s41587-020-0517-0

[19] Hirano H, Gootenberg JS, Horii T, Abudayyeh OO, Kimura M, Hsu PD, et al. Structure and engineering of *Francisella novicida* Cas9. *Cell*. 2016;**164**:950-961. DOI: 10.1016/j.cell.2016.01.039

[20] Pardee K, Green AA, Takahashi MK, Braff D, Lambert G, Lee JW, et al. Rapid, low-cost detection of Zika Virus using programmable biomolecular components. *Cell*. 2016;**165**:1255-1266. DOI: 10.1016/j.cell.2016.04.059

[21] Auslander S, Fussenegger M. Toehold gene switches make big footprints. *Nature*. 2014;**516**:333-334. DOI: 10.1038/516333a

[22] Zhou W, Hu L, Ying L, Zhao Z, Chu PK, Yu XF. A CRISPR-Cas9-triggered strand displacement amplification method for ultrasensitive DNA detection. *Nature Communications*. 2018;**9**:5012. DOI: 10.1038/s41467-018-07324-5

[23] Huang M, Zhou X, Wang H, Xing D. Clustered regularly interspaced short palindromic repeats/Cas9 triggered isothermal amplification for site-specific nucleic acid selection. *Analytical*

Chemistry. 2018;**90**:2193-2200. DOI: 10.1021/acs.analchem.7b04542

[24] Wang X, Xiong E, Tian T, Cheng M, Lin W, Sun J, et al. CASLFA: CRISPR/Cas9-mediated lateral flow nucleic acid assay. *ACS Nano*. 2020;**14**:2497-2508. DOI: 10.1021/acsnano.0c00022

[25] Azhar M, Phutela R, Kumar M, Ansari AH, Rauthan R, Gulati S, et al. Rapid and accurate nucleobase detection using FnCas9 and its application in COVID-19 diagnosis. *Biosensors & Bioelectronics*. 2021;**183**:113207. DOI: 10.1016/j.bios.2021.113207

[26] Quan J, Langelier C, Kuchta A, Batson J, Teyssier N, Lyden A, et al. FLASH: A next-generation CRISPR diagnostic for multiplexed detection of antimicrobial resistance sequences. *Nucleic Acids Research*. 2019;**47**:e83. DOI: 10.1093/nar/gkz418

[27] Bray F, Laversanne M, Weiderpass E, Soerjomataram I. The ever-increasing importance of cancer as a leading cause of premature death worldwide. *Cancer*. 2021;**127**:3029-3030. DOI: 10.1002/cncr.33587

[28] Hajian R, Balderston S, Tran T, de Boer T, Etienne J, Sandhu M, et al. Detection of unamplified target genes via CRISPR-Cas9 immobilized on a graphene field-effect transistor. *Nature Biomedical Engineering*. 2019;**3**:427-437

[29] Louis DN, Perry A, Wesseling P, Brat DJ, Cree IA, Figarella-Branger D, et al. The 2021 WHO Classification of Tumors of the Central Nervous System: A summary. *Neuro-Oncology*. 2021;**23**:1231-1251. DOI: 10.1093/neuonc/noab106

[30] Yang H, Ye D, Guan KL, Xiong Y. IDH1 and IDH2 mutations in tumorigenesis: Mechanistic insights

and clinical perspectives. *Clinical Cancer Research*. 2012;**18**:5562-5571. DOI: 10.1158/1078-0432.ccr-12-1773

[31] Powter B, Jeffreys SA, Sareen H, Cooper A, Brungs D, Po J, et al. Human TERT promoter mutations as a prognostic biomarker in glioma. *Journal of Cancer Research and Clinical Oncology*. 2021;**147**:1007-1017. DOI: 10.1007/s00432-021-03536-3

[32] Buczkowicz P, Hoeman C, Rakopoulos P, Pajovic S, Letourneau L, Dzamba M, et al. Genomic analysis of diffuse intrinsic pontine gliomas identifies three molecular subgroups and recurrent activating ACVR1 mutations. *Nature Genetics*. 2014;**46**:451-456. DOI: 10.1038/ng.2936

[33] Wu G, Broniscer A, McEachron TA, Lu C, Paugh BS, Becksfors J, et al. Somatic histone H3 alterations in paediatric diffuse intrinsic pontine gliomas and non-brainstem glioblastomas. *Nature Genetics*. 2012;**44**:251-253. DOI: 10.1038/ng.1102

[34] Pratilas CA, Xing F, Solit DB. Targeting oncogenic braf in human cancer. *Current Topics in Microbiology and Immunology*. 2012;**355**:83. DOI: 10.1007/82_2011_162

[35] Segerman B. The most frequently used sequencing technologies and assembly methods in different time segments of the bacterial surveillance and Refseq genome databases. *Frontiers in Cellular and Infection Microbiology*. 2020;**10**:527102

[36] Allali I, Arnold JW, Roach J, Cadenas MB, Butz N, Hassan HM, et al. A comparison of sequencing platforms and bioinformatics pipelines for compositional analysis of the gut microbiome. *BMC Microbiology*. 2017;**17**:194. DOI: 10.1186/s12866-017-1101-8

[37] Wang Y, Zhao Y, Bollas A, Wang Y, Au KF. Nanopore sequencing technology, bioinformatics and applications. *Nature Biotechnology*. 2021;**21**:1348

[38] Nicholls SM, Quick JC, Tang S, Loman NJ. Ultra-deep, long-read nanopore sequencing of mock microbial community standards. *Gigascience*. 2019;**8**:giz043

[39] Ardui S, Ameer A, Vermeesch JR, Hestand MS. Single molecule real-time (SMRT) sequencing comes of age: Applications and utilities for medical diagnostics. *Nucleic Acids Research*. 2018;**46**:2159-2168. DOI: 10.1093/nar/gky066

[40] Chen Z, He X. Application of third-generation sequencing in cancer research. *Medical Review*. 2021;**1**:150-171. DOI: 10.1515/mr-2021-0013

[41] Gouil Q, Keniry A. Latest techniques to study DNA methylation. *Essays in Biochemistry*. 2019;**63**:639-648

[42] Usui G, Matsusaka K, Mano Y, Urabe M, Funata S, Fukayama M, et al. DNA methylation and genetic aberrations in gastric cancer. *Digestion*. 2021;**102**:25-32. DOI: 10.1159/000511243

[43] Cao J, Yan Q. Cancer epigenetics, tumor immunity, and immunotherapy. *Trends in Cancer*. 2020;**6**:580-592. DOI: 10.1016/j.trecan.2020.02.003

[44] Villanueva L, Álvarez-Errico D, Esteller M. The contribution of epigenetics to cancer immunotherapy. *Trends in Immunology*. 2020;**41**:676-691. DOI: 10.1016/j.it.2020.06.002

[45] Li S, Tollefsbol TO. DNA methylation methods: Global DNA methylation and methylomic analyses. *Methods*. 2021;**187**:28-43. DOI: 10.1016/j.ymeth.2020.10.002

- [46] Thudi M, Li Y, Jackson SA, May GD, Varshney RK. Current state-of-art of sequencing technologies for plant genomics research. *Briefings in Functional Genomics*. 2012;1:3-11. DOI: 10.1093/bfpgp/elr045
- [47] Amarasinghe SL, Su S, Dong X, Zappia L, Ritchie ME, Gouil Q. Opportunities and challenges in long-read sequencing data analysis. *Genome Biology*. 2020;21:30. DOI: 10.1186/s13059-020-1935-5
- [48] Buytaers FE, Saltykova A, Denayer S, Verhaegen B, Vanneste K, Roosens NHC, et al. Towards real-time and affordable strain-level metagenomics-based foodborne outbreak investigations using oxford nanopore sequencing technologies. *Frontiers in Microbiology*. 2021;21:738284. DOI: 10.3389/fmicb.2021.738284
- [49] Ebbert MTW, Farrugia SL, Sens JP, Jansen-West K, Gendron TF, Prudencio M, et al. Long-read sequencing across the C9orf72 “GGGGCC” repeat expansion: Implications for clinical use and genetic discovery efforts in human disease. *Molecular Neurodegeneration*. 2018;13:46. DOI: 10.1186/s13024-018-0274-4
- [50] Gilpatrick T, Lee I, Graham JE, Raimondeau E, Bowen R, Heron A, et al. Targeted nanopore sequencing with Cas9-guided adapter ligation. *Nature Biotechnology*. 2020;38:433-438. DOI: 10.1038/s41587-020-0407-5
- [51] Li H. Minimap2: Pairwise alignment for nucleotide sequences. *Bioinformatics*. 2018;34:3094-3100. DOI: 10.1093/bioinformatics/bty191
- [52] Cone EB, Marchese M, Paciotti M, Nguyen DD, Nabi J, Cole AP, et al. Assessment of time-to-treatment initiation and survival in a cohort of patients with common cancers. *JAMA Network Open*. 2020;3:e2030072. DOI: 10.1001/jamanetworkopen.2020.30072
- [53] Gerson SL. MGMT: Its role in cancer aetiology and cancer therapeutics. *Nature Reviews. Cancer*. 2004;4:296-307. DOI: 10.1038/nrc1319
- [54] Fey A, Eichler S, Flavier S, Christen R, Höfle MG, Guzmán CA. Establishment of a real-time PCR-based approach for accurate quantification of bacterial RNA targets in water, using *Salmonella* as a model organism. *Applied and Environmental Microbiology*. 2004;70:3618-3623. DOI: 10.1128/aem.70.6.3618-3623.2004
- [55] Garcia HH, Castillo Y, Gonzales I, Bustos JA, Saavedra H, Jacob L, et al. Low sensitivity and frequent cross-reactions in commercially available antibody-detection ELISA assays for *Taenia solium* cysticercosis. *Tropical Medicine & International Health*. 2018;23:101-105. DOI: 10.1111/tmi.13010
- [56] Rang FJ, Kloosterman WP, de Ridder J. From squiggle to basepair: Computational approaches for improving nanopore sequencing read accuracy. *Genome Biology*. 2018;19:1-11. DOI: 10.1186/s13059-018-1462-9
- [57] Grada A, Weinbrecht K. Next-generation sequencing: Methodology and application. *The Journal of Investigative Dermatology*. 2013;133:1-4. DOI: 10.1038/jid.2013.248
- [58] Lin Y, Cradick TJ, Brown MT, Deshmukh H, Ranjan P, Sarode N, et al. CRISPR/Cas9 systems have off-target activity with insertions or deletions between target DNA and guide RNA sequences. *Nucleic Acids Research*. 2014;42:7473
- [59] Tiscione NB. The validation of ELISA screening according to SWGTOX

recommendations. *Journal of Analytical Toxicology*. 2018;**42**:e33-e34

[60] Luo C, Tsementzi D, Kyrpides N, Read T, Konstantinidis KT. Direct comparisons of Illumina vs. Roche 454 sequencing technologies on the same microbial community DNA sample. *PLoS ONE*. 2012;**7**:2

[61] Roberts RJ, Carneiro MO, Schatz MC. The advantages of SMRT sequencing. *Genome Biology*. 2013;**14**:1-4

[62] Euskirchen P, Bielle F, Labreche K, Kloosterman WP, Rosenberg S, Daniau M, et al. Same-day genomic and epigenomic diagnosis of brain tumors using real-time nanopore sequencing. *Acta Neuropathologica*. 2017;**134**:691-703. DOI: 10.1007/s00401-017-1743-5

[63] Weerasinghe RK, Meng R, Dowdell AK, Bapat B, Vita A, Schroeder B, et al. Identification of clinically actionable biomarkers via routine comprehensive genomic profiling across a large community health system. *Journal of Clinical Oncology*. 2022;**40**:e15035. DOI: 10.1200/JCO.2022.40.16_suppl.e15035

[64] Seufi AEM, Galal FH. View of fast DNA purification methods: Comparative study. *WAS Science Nature*. 2020;**3**. Available from: <http://worldscience.org/journals/index.php/wassn/article/view/9>