Molecular Aspects of Medicine 96 (2024) 101253

Contents lists available at ScienceDirect

Molecular Aspects of Medicine

journal homepage: www.elsevier.com/locate/mam

Review

Principles of digital sequencing using unique molecular identifiers

Daniel Andersson [a], Firaol Tamiru Kebede [a], Mandy Escobar [a], Tobias Österlund [a,b,c], Anders Ståhlberg [a,b,c,*]

[a] Sahlgrenska Center for Cancer Research, Department of Laboratory Medicine, Institute of Biomedicine, Sahlgrenska Academy, University of Gothenburg, 413 90, Gothenburg, Sweden
[b] Wallenberg Centre for Molecular and Translational Medicine, University of Gothenburg, 413 90, Gothenburg, Sweden
[c] Department of Clinical Genetics and Genomics, Sahlgrenska University Hospital, 413 45, Gothenburg, Sweden

ARTICLE INFO

Keywords:
Digital sequencing
Molecular barcode
Ultrasensitive sequencing
Unique molecular identifier

ABSTRACT

Massively parallel sequencing technologies have long been used in both basic research and clinical routine. The recent introduction of digital sequencing has made previously challenging applications possible by significantly improving sensitivity and specificity to now allow detection of rare sequence variants, even at single molecule level. Digital sequencing utilizes unique molecular identifiers (UMIs) to minimize sequencing-induced errors and quantification biases. Here, we discuss the principles of UMIs and how they are used in digital sequencing. We outline the properties of different UMI types and the consequences of various UMI approaches in relation to experimental protocols and bioinformatics. Finally, we describe how digital sequencing can be applied in specific research fields, focusing on cancer management where it can be used in screening of asymptomatic individuals, diagnosis, treatment prediction, prognostication, monitoring treatment efficacy and early detection of treatment resistance as well as relapse.

1. Introduction

Our understanding of human physiology and pathological conditions has advanced substantially the last few decades due to development of methods that enable analysis of nucleic acids. Two of the most commonly used technologies applied in basic research and clinical routine are quantitative PCR and massively parallel sequencing. Quantitative PCR is simple and cost-effective, but analysis is limited to a few target sequences. In contrast, sequencing has an advanced experimental workflow that enables analysis of entire genomes but at a higher cost. Both PCR- and sequencing-based methods allow detection of individual molecules. However, they lack the sensitivity to detect rare sequence variants in large backgrounds, being limited to detect variant allele frequencies higher than ~1% due to polymerase-induced errors introduced during amplification and sequencing (Fox et al., 2014; Stead et al., 2013; Xu, H. et al., 2014). In several emerging applications, such as circulating tumor-DNA analysis (Cescon et al., 2020) and immune repertoire profiling (Johansson et al., 2020), a sensitivity to assess variant allele frequencies <0.1% is often crucial for optimal clinical interpretation. This issue was addressed with the introduction of digital PCR and digital sequencing that were developed to facilitate robust detection of low variant allele frequencies, even the detection of individual molecules with specific nucleotide variants, *i.e.*, digital approach. Like quantitative PCR, digital PCR is only suitable for analysis of single or few target sequences. Conversely, digital sequencing enables assessment of far larger numbers of target sequences. Besides superior sensitivity, digital approaches also provide more accurate quantification of target molecules, since they minimize technical biases that are associated with conventional PCR and sequencing. The pros and cons of the different methods are illustrated in Fig. 1. In this review, we outline the principles of digital sequencing, discussing its potentials and limitations and illustrate how it can be used with examples from applications in cancer management.

1.1. Types of unique molecular identifiers

Digital sequencing utilizes unique molecular identifiers (UMIs), also known as molecular barcodes, to minimize the technical noise and biases inherent to sequencing. The concept of UMIs was introduced two decades ago (Hug and Schuler, 2003) and has been experimentally implemented along with the development of new sequencing approaches. The UMI is a random DNA sequence that labels each template

* Corresponding author. Sahlgrenska Center for Cancer Research, University of Gothenburg, Box 425, 405 30, Gothenburg, Sweden.
E-mail address: anders.stahlberg@gu.se (A. Ståhlberg).

https://doi.org/10.1016/j.mam.2024.101253
Received 16 November 2023; Received in revised form 26 January 2024; Accepted 3 February 2024
Available online 16 February 2024

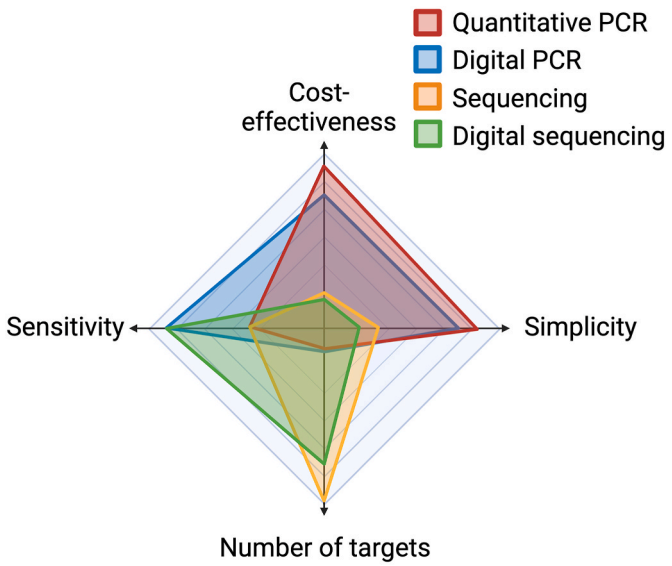

Fig. 1. Basic characteristics of PCR- and sequencing based methods. The relative level of sensitivity, cost-effectiveness and simplicity as well as relative number of target sequences are shown in relation to the performance of quantitative PCR, digital PCR, sequencing and digital sequencing. Image created with BioRender.com.

DNA molecule of interest (Fig. 2A). During library construction, the UMI is amplified alongside the targeted sequences. Thus, following sequencing, all sequence reads with identical UMI, *i.e.*, members of the same UMI family, originate from the same template DNA molecule. Bioinformatically, all reads in a UMI family are collapsed into a consensus read that enables both correction of polymerase-induced errors as well as minimization of quantification biases, also known as PCR duplicates. As the removal of technical polymerase-induced errors, such as single nucleotide variants, inserts and deletions, relies on consensus reads containing several members, the correction of polymerase-induced errors is only possible when each original DNA molecule has been sequenced multiple times.

The UMI may already exist in template DNA, *i.e.*, endogenous UMI, or can be introduced experimentally, *i.e.*, exogenous UMI. Digital sequencing may use either one or both types of UMI in combination. Different aspects of endogenous UMIs can be used. One alternative is to exploit that the template DNA is naturally fragmented within the body during and after cell death (Bronkhorst et al., 2019; Heitzer et al., 2020). For instance, cell-free DNA in blood is rapidly degraded, where nucleosome-protected DNA remains intact for a longer time than DNA that is not associated with histone complexes. DNA fragmentation can also be artificially performed before library preparation. Regardless of how fragmentation is attained, DNA breaks will occur at different locations and result in fragments with varying start and stop positions as well as sizes. Another subtype of endogenous UMIs is genomic sequences that are differentially recombined among cells. V(D)J recombination in B and T cells are somatic variations that give rise to unique clones, where each individual rearrangement can be viewed as a UMI (Bolotin et al., 2015; Johansson et al., 2020; Wu, D. et al., 2012). Endogenous UMIs can also be experimentally introduced by converting unmethylated cytosine nucleotides to uracil with a bisulfite conversion step before library construction (Mattox et al., 2017). Exogenous UMIs typically consist of 6–12 nucleotides long randomized sequences that are experimentally attached to target DNA at either the 5′- or 3′-end, alternatively at both ends. Several additional concepts exist to utilize endogenous and exogenous UMIs, either individually or in combination.

1.2. Properties of unique molecular identifiers

To enable sequencing error correction, the pool of specific UMIs, *i.e.*, the UMI diversity, needs to be substantially larger than the number of analyzed DNA molecules to reduce the risk of UMI collisions, *i.e.*, the probability that two or more template molecules are labeled with identical UMIs. The diversity of endogenous UMIs is challenging to estimate since DNA fragmentation as well as DNA recombination are both dependent on processes that are not completely arbitrary. In comparison, the theoretical diversity of exogenous UMIs can easily be calculated from the UMI length and the number of nucleotide types each given position is allowed to have, where Diversity $= 4^{\text{number of fully randomized nucleotides}}$. For example, to reach a diversity over 10^6, ten fully randomized nucleotides are needed. In contrast, if only two types of nucleotide bases are allowed, twice the UMI length is required to achieve the same diversity. If the UMIs are attached to both ends of the DNA strand, the diversity is calculated from the total sum of all nucleotides present at both UMI segments. However, the practical diversity of exogenous UMI is normally considerably lower than the theoretical. This is partly because all oligonucleotide variants are not generated to the same extent during chemical oligonucleotide synthesis (Filges et al., 2021). While long UMIs offer higher diversity than short, they are experimentally more challenging to work with since random sequences are prone to interact with each other, generating unspecific library products that may consume significant sequencing resources or even cause library construction and sequencing to fail altogether, especially in samples with limited amount of DNA (Ståhlberg et al., 2016). Another drawback, although negligible in most applications, is that longer UMIs requires additional cycles during sequencing to read through the UMI itself. Thus, the UMI length should be selected based on application and to avoid the risk of UMI collisions (Fig. 2B and C). For instance, the assessment of cell-free DNA in liquid biopsies is often limited to 5–40 ng DNA per sample, and thereby do not generally need as high UMI diversity as analysis of most standard tissue biopsies that provide several micrograms of DNA.

While digital sequencing corrects for sequencing errors in the DNA of interest, it cannot correct for polymerase-induced errors in the UMI sequence itself during library generation and sequencing, which ultimately will result in false UMIs. Examples of this are imbalanced GC-content and homopolymer sequences that are known to perform poorly in sequencing. To overcome these issues, the UMI designs may include nucleotides that are not fully randomized to eliminate sequences that are especially prone to give rise to errors (Johnson et al., 2023; Karst et al., 2021). The drawback is that these strategies result in longer UMIs to maintain diversity. Overall, the optimal UMI design is poorly studied using experimental data.

1.3. Unique molecular identifiers enable digital sequencing

There is a continuous increase of digital sequencing approaches (Table 1). They utilize different experimental protocols to suit specific needs, ranging from targeting individual DNA sequences to entire genomes (Salk et al., 2018). Most digital sequencing methods are either PCR- or ligation-based and utilize exogenous UMIs. Regardless method, the UMI should label target DNA as early as possible in the library construction to minimize sequencing errors. In PCR-based methods, UMIs are incorporated as a part of the primers that target DNA sequences of interest. Conversely, in ligation-based approaches all DNA is labeled with adapters containing UMIs by a ligation step, where sequences of interest often are enriched by a subsequent hybridization-capture step. Fig. 3 shows the experimental principles of SiMSen-Seq and DupSeq, as examples of PCR- and ligation-based approaches, respectively. The labeling with exogenous UMIs can occur on either single-stranded or double-stranded DNA molecules, where the latter allows for higher confidence in variant calling as polymerase-induced errors are unlikely to appear in both DNA strands at

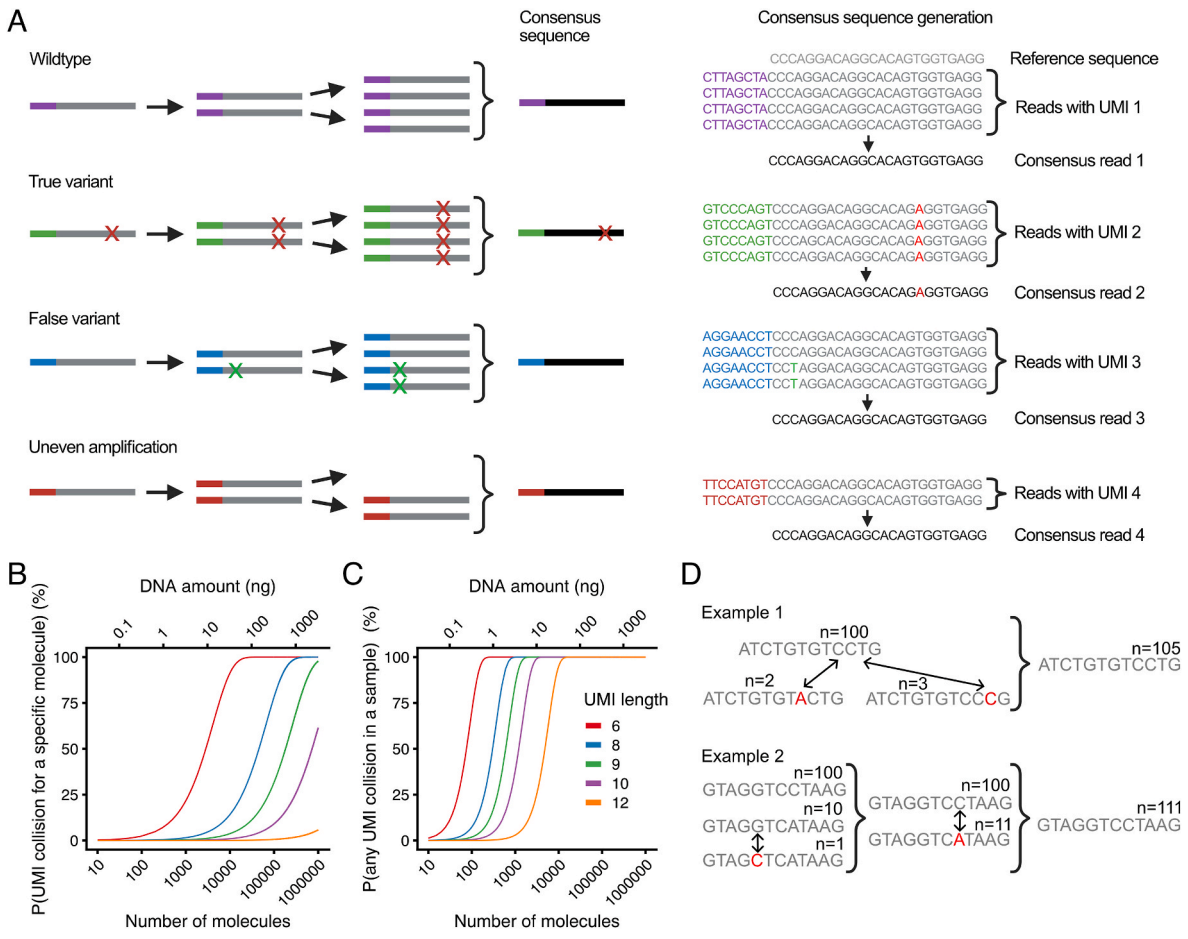

Fig. 2. The concept and properties of UMIs in digital sequencing. (A) A conceptual overview of four identical target DNA sequences (gray) labeled with exclusive UMIs at respective 5′-end before amplification (purple, green, blue and red). An example with specific nucleotide sequences after sequencing is shown to the right. Reads are bioinformatically collapsed into consensus reads based on their UMI. The true variant (indicated by a red cross or letter) is present in all amplified molecules, while the false variant (green cross or letter) is only present in a subset of amplified molecules. (B) The probability of UMI collisions. The probability that a specific DNA molecule will be labeled by a UMI identical to that of another DNA molecule is shown in relation to different UMI lengths and number of molecules, where 1 ng of human DNA corresponds to approximately 310 haploid genome equivalents (Piovesan et al., 2019). The probability of UMI collision is calculated as $1 - \left(\frac{n-1}{n}\right)^k$, where n is the number of possible UMIs ($4^{\text{number of fully randomized nucleotides}}$) and k is the number of molecules. This risk of DNA collision is neglectable for UMI lengths $\geq$10 when analysis is restricted to a few thousand molecules. (C) The probability that any two or more DNA molecules in the sample will be labeled with identical UMI as a function of UMI lengths and number of molecules. The probability of any UMI collision is approximated with the formula $1 - e^{\frac{-k^2}{2n}}$, where n is the number of possible UMIs ($4^{\text{number of fully randomized nucleotides}}$) and k is the number of molecules (Flajolet et al., 1992). For example, in a sample with in total 1000 molecules, the risk that two molecules are labeled with identical UMI is 99.95%, 37.9% and 2.9%, using UMI length 8, 10 and 12, respectively. Note that the risk of UMI collision is relatively high for most conditions. However, the practical impact of this is limited for most digital sequencing applications since they generally aim to detect rare or even individual molecules, making the case scenario in subplot B with a single molecule more relevant. (D) Bioinformatical UMI clustering. There are several strategies to cluster UMIs. Most algorithms use an edit distance of one, *i.e.*, the maximum allowed difference between two connected UMIs is at most one mismatch or insertion/deletion (Smith et al., 2017). The most frequent UMI variant is referred to as a centroid and it is to this other less frequent UMIs are collapsed into. The figure shows two different examples of UMI clusters where the edit distance threshold is set to a value of one. The number of each UMI variant is shown above each sequence.

the same position. Some methods also attach multiple UMIs per original DNA fragment to simplify the experimental workflow (Cohen et al., 2018; Newman et al., 2016; Sagitov and Ståhlberg, 2023; Ståhlberg et al., 2017). A majority of digital sequencing protocols also utilize high-fidelity DNA polymerases in library construction to further reduce errors in target DNA sequences as well as in the UMI sequence. It should be noted that the use of UMIs is far superior to that of any high-fidelity DNA polymerase to minimize sequencing errors in library construction and sequencing (Filges et al., 2019). The use of UMIs allows elimination of almost all sequencing errors, but no digital sequencing method can correct for technical errors that may occur before sequencing, *e.g.*, errors that are introduced during sample ageing, tissue fixation (Do and Dobrovic, 2015; Zimmermann et al., 2008) and other pre-analytical steps (Chen et al., 2017; Costello et al., 2013).

Besides choice of UMI type, there are several other factors that need to be taken into consideration when choosing optimal digital sequencing approach. Among these are DNA quantity and quality, size of target DNA panel to be assessed, sensitivity and specificity to detect low variant allele frequencies, simplicity of workflow, analytical turnaround time and cost per sample. PCR-based methods tend to have less complex and time-consuming experimental workflow than ligation-based methods, especially if hybridization-capture is used to enrich target sequences. However, the effort to develop *de novo* PCR-based gene panels increases with the total size of target DNA sequences. Hence, PCR-based approaches are mostly used for small to medium sized gene panels (ranging from 10^2 to 10^5 target nucleotides), whereas a ligation-based method is the preferred choice for medium to large gene panels (ranging from 10^4 to 10^6 target nucleotides). Depending on sample properties, DNA

Table 1
Examples of digital sequencing approaches.

UMI type[a]	Method	Sequencing approach	Reference
Ligation based[b]			
Exogenous	CODEC	Targeted WGS	Bae et al., 2023
Exogenous	MAESTRO	Targeted	Gydush et al., 2022
Exogenous	SSM-Seq	WGS	Maslov et al., 2022
Exogenous	NanoSeq	WGS Targeted	Abascal et al., 2021
Exogenous	SaferSeqS	Targeted	Cohen et al., 2021
Exogenous	SLHC-Seq	Targeted	Liu et al., 2019
Exogenous	TARDIS	Targeted	McDonald et al., 2019
Exogenous	cfDNA-Seq	Targeted	Mansukhani et al., 2018
Exogenous	DEEPER-Seq	Targeted	Wang et al., 2017
Exogenous	TrUMIseq	WGS Targeted	Hong and Gresham, 2017
Exogenous	CypherSeq	WGS Targeted	Gregory et al., 2016
Exogenous	NOIR	Targeted	Kukita et al., 2015
Exogenous	DupSeq	WGS Targeted	Kennedy et al., 2014; Schmitt et al., 2012, 2015
Exogenous & Endogenous	LTC	Targeted	Pel et al., 2018
Exogenous & Endogenous	TEC-Seq	Targeted	Phallen et al., 2017
Exogenous & Endogenous	iDES	Targeted	Newman et al., 2016
Exogenous & Endogenous	BotSeqS	WGS	Hoang et al., 2016
Endogenous	Safe-SeqS	Targeted	Kinde et al., 2011
Endogenous	Subassembly	WGS	Hiatt et al., 2010
PCR-based[b]			
Exogenous	SiMSen-Seq	Targeted	Ståhlberg et al., 2016
Exogenous	MIPSTR	Targeted	Carlson et al., 2015
Exogenous	smMIP	Targeted	Hiatt et al., 2013
Exogenous	Narayan et al. (2012)	Targeted	Narayan et al., 2012
Exogenous	Safe-SeqS	Targeted	Kinde et al., 2011
Exogenous	BiSeqS	Targeted	Mattox et al., 2017

WGS, whole genome sequencing.
[a] UMI type indicated in original study is shown.
[b] Approach to incorporate UMI in original study is shown.

quantity and quality may vary substantially, where not all digital sequencing approaches are equally suitable. Freshly collected and processed samples from cell cultures, blood buffy coats and tissues often yield large amounts of intact DNA. In comparison, extractions from cell-free liquid biopsies, forensic and environmental samples generally result in minor DNA amounts that are also often highly fragmented. Some sample types, moreover, regularly contain inhibitors that affect downstream enzymatic reactions, *e.g.*, library construction (Sidstedt et al., 2020). High sensitivity refers to the ability to detect true variants, *i.e.*, identification of molecules with variant sequences in a wildtype background. Importantly, though often overlooked, sensitivity is also associated with the ability to find all original template molecules in the sample. Here, methods that efficiently separate true variants from wildtype background but only manage to assess small fractions of all original molecules are useful in applications with high DNA yields but not in applications where the expected DNA yield is limited. High specificity refers to the ability to avoid classifying technical noise as biological variants.

Another emerging field that utilizes digital sequencing is single-cell analysis, where RNA profiling is the most developed. There are several single-cell technology reviews highlighting the potential and limitations of various methods in relation to specific applications (Baysoy et al., 2023; Heumos et al., 2023; Hwang et al., 2018; Jovic et al., 2022; Luecken and Theis, 2019). There are two main strategies for quantification of the entire transcriptome. Methods like Smart-seq3 can assess the full-length transcripts with strand specificity (Hagemann-Jensen et al., 2020), while droplet-based approaches only detect parts of the transcripts with the aim to determine their expression levels (Zheng et al., 2017). Droplet-based single-cell technologies enable high-throughput analysis of large cell numbers, while whole transcript analysis is more suited for smaller cell numbers where information about isoforms and allele variants is required. In droplet-based technologies, UMIs are combined with cellular barcodes that also consist of a random DNA sequence. While UMIs separate individual nucleic acid molecules from each other, cellular barcodes separate cells from each other and are used as a cell index to link all sequencing reads from a given cell together. Despite different uses, the two barcode types are experimentally and bioinformatically handled similarly.

In conclusion, digital sequencing enables high sensitivity and specificity using UMIs. Fig. 4A illustrates an example where a true variant can be detected only using UMIs where the background noise is otherwise too high. There is a vast diversity of experimental protocols, often tailor-made for specific sample types or applications, with numerous ways of utilizing endogenous and exogenous UMIs to minimize sequencing errors in digital sequencing. However, to systematically evaluate method performances, they need to be compared side-by-side using relevant and standardized reference materials (Stetson et al., 2019; Weber et al., 2020).

1.4. Bioinformatics to handle unique molecular identifiers

Digital sequencing reads need to be processed with UMI-aware bioinformatics workflows. Table 2 shows examples of available bioinformatics pipelines and tools for UMI data processing and analysis. Normally, quality controls and trimming of the reads are first performed, followed by extraction of the UMI sequence from each read. The reads are subsequently aligned to a reference genome and grouped into UMI families, where reads with identical UMI sequence and alignment position to the reference genome will be grouped together, *i.e.*, deduplication. To allow for sequencing errors in the UMI sequence, close-to-identical UMIs can be clustered together (Smith et al., 2017) (Fig. 2D). Here, an edit distance of one is often applied to group UMIs with maximum one nucleotide difference into the same UMI family. An inevitably shortcoming with UMI clustering is that some UMIs may erroneously be grouped. However, UMI clustering is advantageous in most applications. To validate and compare bioinformatics tools, data sets generated from standardized control materials with known concentrations and variant allele frequencies are needed. The reference material should reflect both the complexity of the sample matrix as well as the expected amounts of target molecules, including low variant allele frequencies, for a given application. For example, cell-free DNA standards are typically fragmented with variant allele frequencies between 0.1% and 5%.

For some applications, such as sequencing of short tandem repeats and the hypervariable regions of the 16S and 18S rRNA genes, the alignment step can be skipped. Here, the sequences of target primers or hybridization probes may be used together with the UMI sequence to create UMI families (Shugay et al., 2014). In RNA sequencing, including single-cell RNA sequencing, the UMIs are primarily used to correct for transcript quantification biases caused by PCR duplicates by counting each UMI family, regardless of size, only once. Preprocessing and UMI-counting methods for single cell RNA-sequencing have recently been reviewed and benchmarked (Gao et al., 2021; You et al., 2021).

1.5. Unique molecular identifier-based error correction requires deep sequencing

If the workflow of digital sequencing becomes as simple as for conventional sequencing, should UMIs be used in all applications? To enable UMI-based error correction, deep sequencing is needed where each target DNA molecule is sequenced multiple times. Fig. 4B shows how many times each UMI is sequenced for a typical target, where UMIs

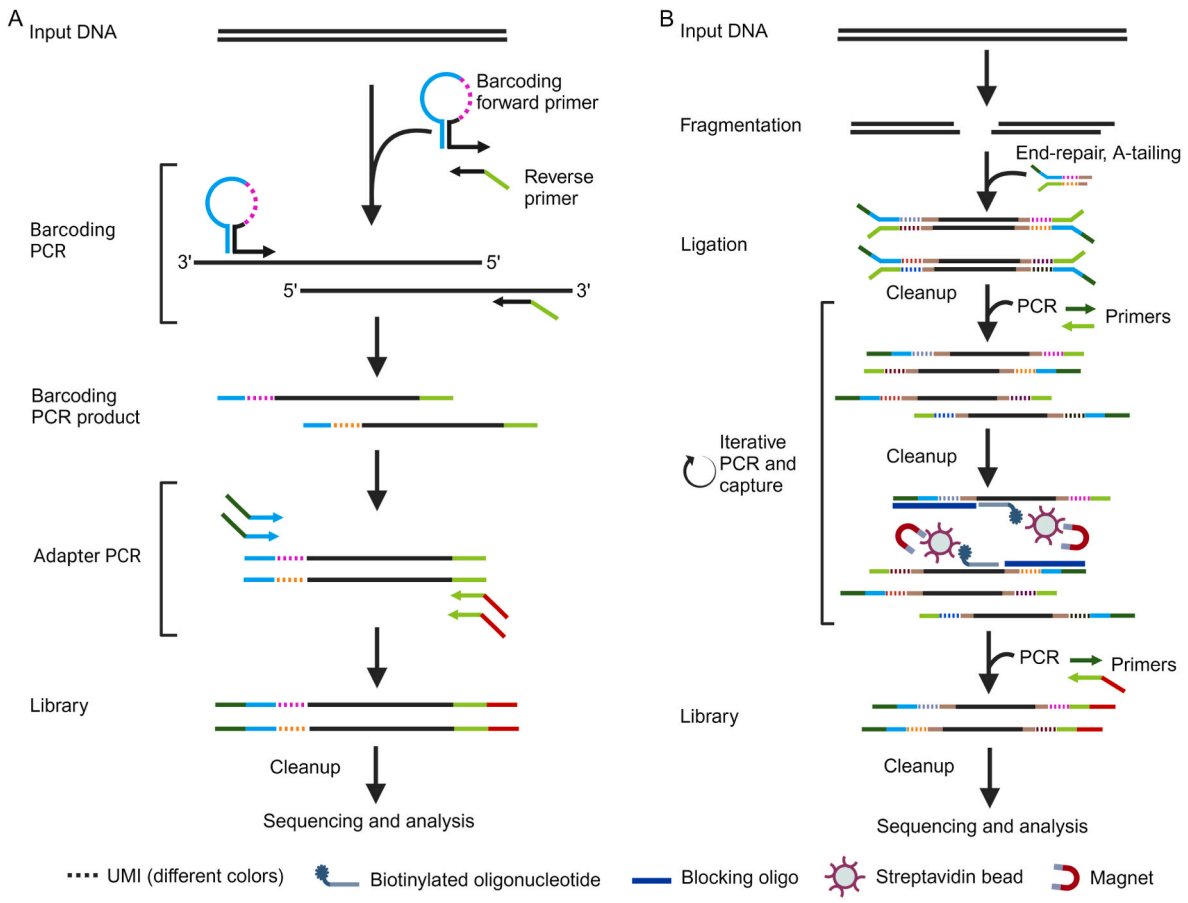

Fig. 3. Digital sequencing approaches. (A) Schematic overview of SiMSen-Seq as an example of PCR-based approach (Ståhlberg et al., 2016). The protocol consists of two sequential PCR steps. In the first, target DNA is labeled with UMIs using PCR primers during three cycles. In the second PCR, UMI-labeled target DNA is amplified with sequencing adapters. Libraries are then purified and sequenced. (B) Schematic overview of DupSeq as an example of ligation-based approach (Schmitt et al., 2015). Hybridization-capture DupSeq protocol starts with fragmentation of double-stranded input DNA, a step that is not needed for degraded DNA sample types, such as cell-free DNA in liquid biopsies. Ends of fragmented DNA are repaired and A-tailed enabling ligation to T-tailed adapter primers containing UMIs. The adapter-ligated DNA is purified with beads and amplified using PCR. The library is again cleaned up and hybridized to a pool of biotinylated DNA oligonucleotides, complementary to all sequences of interest. The hybridized molecules are enriched using magnetic streptavidin-coated beads. The PCR and hybridization-capture step is repeated for smaller panels, while the iteration can be omitted for larger panels like whole exome. Finally, enriched target DNA is amplified with sequencing adapters, purified and sequenced. Image created with BioRender.com.

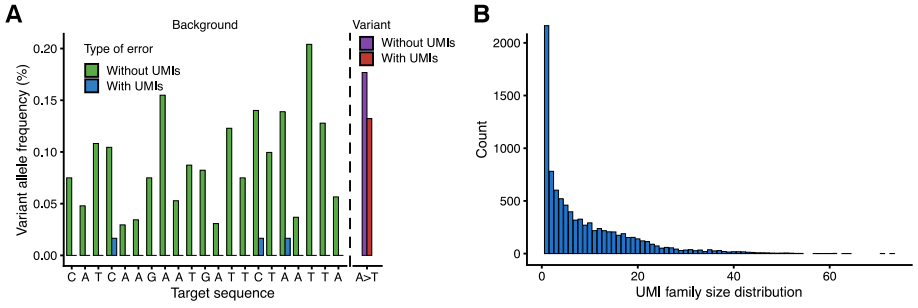

Fig. 4. Digital sequencing enables ultrasensitive variant detection. (A) Experimental data demonstrating the use of UMIs to detect a low frequent variant. A target sequence containing 22 nucleotides was analyzed in a reference sample with a known variant, with and without the use of UMI-based error correction. Data show that the expected variant cannot be distinguished from the other nucleotide positions when the UMIs are not bioinformatically utilized, while the variant is clearly identifiable following UMI-based error correction. For 18 nucleotide positions no errors are observed using UMI-based error correction. (B) Unique molecular identifier family sizes for a given target sequence. The diagram shows the number of times each UMI-labeled molecule is sequenced for a given target sequence. In SiMSen-Seq data analysis, three reads per UMI family is usually applied as cut-off to generate consensus reads. In this example, 2944 UMI families had less than three reads, while 6138 UMI families had three or more reads. Experimental data is from a published data set using SiMSen-Seq (Österlund et al., 2022).

appearing only once by default prohibit error correction in downstream bioinformatics and, conversely, UMIs with numerous reads are sequenced more than needed. The demand of sequencing capacity is proportional to the target panel size and the amount of analyzed DNA (Fig. 5). For example, the human haploid genome consists of ~3.055 billion base-pairs (Nurk et al., 2022) that can be analyzed as 100 base-pairs long sequences, resulting in 30,550,000 DNA sequences. Twenty nanogram DNA consist of 6,200 copies of each DNA sequence

Table 2
Bioinformatics tools for handling sequencing data with UMIs.

Workflow	Tool	Reference
Calib	UMI deduplication/UMI clustering	Orabi et al., 2019
DAUMI	UMI deduplication/UMI clustering	Peng and Dorman, 2023
UMI-tools	UMI deduplication/UMI clustering	Smith et al., 2017
SASCRiP	UMI deduplication, single-cell RNA-seq	Moonsamy and Gentle, 2022
BUStools	UMI deduplication format, single-cell RNA-seq	Melsted et al., 2019
UMIc	UMI deduplication, consensus generation	Tsagiopoulou et al., 2021
DeepSNVMiner	UMI deduplication, variant calling	Andrews et al., 2016
MAGERI	UMI deduplication, variant calling	Shugay et al., 2017
UMIErrorCorrect	UMI deduplication, consensus generation, variant calling	Österlund et al., 2022
MIGEC	UMI deduplication, assembly (alignment-free), immune cell analysis	Shugay et al., 2014
SmCounter2	Variant calling	Xu, C. et al., 2019
UMI-VarCal	Variant calling	Sater et al., 2020

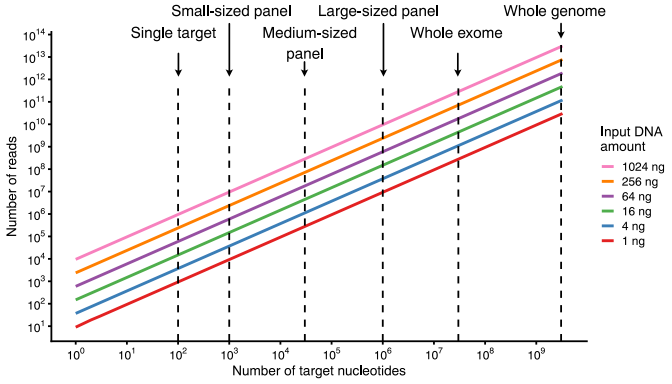

Fig. 5. The number of required reads in relation to total size of target sequence and amount of input DNA using digital sequencing. Here, the targeted sequences are divided into 100 nucleotide long reads. We assume that 1 ng DNA corresponds to 310 molecules (Piovesan et al., 2019) and that each molecule needs to be sequenced three times, *i.e.*, three reads per UMI family. Typical sizes of different panels are indicated at the top, where the sizes of small-, medium- and large-sized panels may vary. The output from current Illumina sequencing platforms ranges from 4 million to 104 billion reads. The exome and genome sizes are about 30 million and 3.055 billion base-pairs, respectively.

(Piovesan et al., 2019). If we assume that three reads per UMI and DNA sequence is needed for UMI-based error correction, a minimum of 5.68×10^{11} reads are needed for each sample. To put this into perspective, the Illumina platform that currently has the highest sequencing capacity, the NovaSeq X Plus, generates up to 1.04×10^{11} paired-end reads. Thus, due to current sequencing capacity and costs, digital sequencing is mainly limited to targeted panels, while whole genome approaches are mainly restricted to shallow coverage. For applications, where UMIs are only used to minimize quantification biases (Fig. 2A), such as single-cell RNA sequencing, no additional sequencing is required in comparison with conventional sequencing. Here, all reads with a UMI are used, even UMI families with only one member. For these types of applications, UMIs are expected to be implemented in most future protocols, since the additional technical challenges associated with UMIs are few and generally neglectable.

1.6. Digital sequencing applications focusing on cancer management

Digital sequencing is applied in numerous clinical applications, such as cancer, infections, neurodegenerative diseases, prenatal testing and transplantation. It can also be used in other fields, *e.g.*, forensics, archaeogenetics and environmental DNA assessments (Table 3). Digital sequencing can be applied to any analyte or biomarker that uses DNA sequences as a readout, such as RNA as well as DNA and RNA interactions with proteins and antibodies. Naturally, the preanalytical and experimental steps prior to the actual DNA library construction step vary substantially between analysis of these various analytes.

In cancer management, digital sequencing can add clinical value in several applications (Fig. 6). Cancer is a heterogeneous group of diseases with different clinical needs regarding biomarker analysis. Cancers

Table 3
Examples of applications with digital sequencing using UMIs.

Area	Application	Reference
Archaeogenetics	Ancient RNA	Mármol-Sánchez et al., 2023
Infectious diseases	Drug resistance	Aydemir et al., 2018; George et al., 2022; Keys et al., 2015
	Virus detection	Ko et al., 2021
Cancer management	Screening	Chabon et al., 2020; Cohen et al., 2018
	Diagnostics	Stackpole et al., 2022; Yang et al., 2019
	Treatment prediction	Suppan et al., 2022
	Prognostication	Barnell et al., 2022; Xu, R. H. et al., 2017; Zhou et al., 2022
	Monitoring of treatment efficacy	Egyud et al., 2019; Johansson et al., 2021
	Treatment response	Hilke et al., 2020; Stankunaite et al., 2022
	Minimal residual disease	Patkar et al., 2021; Ryoo et al., 2023; Tie et al., 2016; Waldeck et al., 2022
	Detection of *de novo* treatment resistance	Varghese et al., 2021; Derrien et al., 2023
	Tumor evolution	Walens et al., 2020; Hu et al., 2021
	Early relapse detection	Qiu et al., 2021
Environmental	Ecosystem health	Fields et al., 2021
Forensics	Mixed and degraded DNA	Jäger et al., 2017; Wu, L. et al., 2019
Prenatal testing		Tam et al., 2020; Zhang et al., 2019
Metagenomics	Resistome	Noyes et al., 2017
Mutagenesis	Genotoxicity evaluation	Valentine et al., 2020; Dodge et al., 2023
Neurodegenerative diseases		Hoekstra et al., 2016; Pandey et al., 2022
Transplantation medicine	Transplant rejection	Kueng et al., 2023
Cellular barcoding	Cell fate mapping, lineage tracing	Biddy et al., 2018; Walens et al., 2020
Metabolic diseases		Yang, S. et al., 2021; Lu et al., 2022
Evolutionary studies		Xia et al., 2020; Zurek et al., 2020
Food science		Smyczynska et al., 2020
DNA sequencing		Kinde et al., 2011; Schmitt et al., 2012
RNA sequencing	Standard RNA sequencing	Fu, C. et al., 2021; Fu, Y. et al., 2018; Hong and Gresham, 2017; Ye et al., 2018
	Single-cell RNA sequencing	Zheng et al., 2017; Pandey et al., 2022
Protein-RNA interactions	iCLIP, eCLIP, sCLIP	Kargapolova et al., 2017; König et al., 2010; Van Nostrand et al., 2016
ChIP seq	AutoRELACS, MINUTE-ChIP	Arrigoni et al., 2020; Kumar and Elsässer, 2019
RNA modifications		Koh et al., 2019
DNA modifications	UMI-ATAC-seq	Zhu et al., 2020
	DNA methylation	Stackpole et al., 2022; Xu, R. H. et al., 2017
Immunosequencing	SiMSen-Seq	Johansson et al., 2020
Antibody detection		Levin et al., 2023

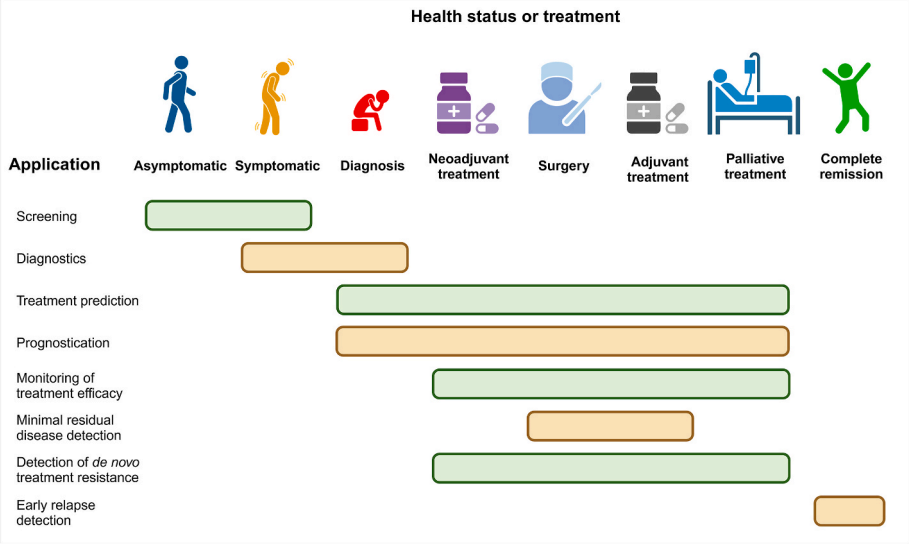

Fig. 6. Digital sequencing applications in cancer management. The timeline for a patient with cancer is shown at the top, where palliative treatment and complete remission stand as alternative outcomes. Different applications and at what stages they can be used and where digital sequencing may add value to clinical decisions are indicated below. Image created with BioRender.com.

display highly variable mutational landscape and expression of biomarkers, not only between but also within specific entities. Some cancers are genetically characterized by recurrent mutations in specific genes and high mutational burden, whereas others have low mutational burden and few hotspot mutations. Tissue biopsies are the golden standard in cancer diagnostics, offering both morphological and molecular information. However, in some settings, tumor biopsies are not feasible to collect. Here, minimally invasive liquid biopsies are emerging as a clinically relevant complement. Unlike tissue biopsies, liquid biopsies can be collected even when no tumor tissue is evident, such as in screening of asymptomatic individuals. Maybe most importantly, most liquid biopsies are also available for repeated sampling, enabling molecular monitoring of treatment efficacy and disease progression. Liquid biopsies can be sampled from in principle any bodily fluid, such as urine, saliva and cerebrospinal fluid, with blood plasma being by far the most commonly used (Andersson et al., 2020). Tumor-derived DNA and other analytes are released from tumor cells into surrounding body fluids via apoptosis, necrosis or active secretion (Bronkhorst et al., 2019). The amount of cell-free DNA in plasma is normally low, between 1 and 10 ng per milliliter blood plasma in healthy individuals (Mead et al., 2011; Mouliere et al., 2014), heavily fragmented to an average size of 167 base-pairs (Ivanov et al., 2015; Snyder et al., 2016) and with a half-life estimated between 15 min and ~2.5 h (Lo et al., 1999; To et al., 2003). The total amount and quality of tumor-DNA are thus comparably much lower in blood plasma than in tissue, where the number of circulating tumor-DNA molecules often is lower than one in a thousand. Digital sequencing is thereby required for most liquid biopsy-based applications.

To cure cancer, it is important to detect the tumor at an early stage before tumor cells invade surrounding tissue or form metastases. To screen asymptomatic individuals for cancer, the use of blood plasma and digital sequencing is an emerging field of research with high potential clinical value. Most attempts have historically been focused on detection of circulating tumor-DNA by assessing hotspot mutations (Chung et al., 2019). An obstacle is that detected DNA variants, even in known cancer driving genes, are not always related to any ongoing malignant disease, *i.e.*, false clinical positives. Clonal hematopoiesis contributes with mutated DNA molecules in blood plasma (Razavi et al., 2019). For example, *TP53* mutations increase with age in otherwise healthy individuals (Salk et al., 2019). Another challenge with mutation analysis in plasma is that many commonly mutated genes, such as *TP53* and *KRAS,* are not tumor type-specific and can therefore be associated to several entities. One strategy to overcome this is to use DNA methylation profiles instead of mutations since DNA methylation patterns are cell type-specific (Moss et al., 2018; Varley et al., 2013). The use of multi-analyte strategies, such as mutations combined with proteins, may also improve clinical sensitivity and specificity (Cohen et al., 2018). For screening applications using mutational analysis, digital sequencing methods that utilize large hotspot panels are preferred to increase the likelihood that one or several tumor-specific mutations can be detected. For high-risk individuals, such as persons with germline mutations in cancer-associated genes like *BRCA1* and *BRCA2*, panels can be smaller since they can be focused to cover mutations only related to a specific diagnosis or a restricted number of diagnoses. The size of a suitable panel depends on tumor type, where some entities, like many sarcomas and pediatric cancers, have few recurrent mutations (Gröbner et al., 2018; Taylor et al., 2011; Vogelstein et al., 2013). Here, large panels may not be enough, instead requiring whole exome or even whole genome sequencing.

In diagnostics, digital sequencing may be useful when the number of tumor cells are very few in relation to non-neoplastic cells, even in tissue biopsies. This occurs when the tumor cell frequency is at the same sensitivity level as standard clinical sequencing, normally 1%. For some tumor types it may be important not to sample the actual tumor tissue since the sampling procedure may disseminate tumor cells (Ghiam et al., 2019). Digital sequencing approaches and using liquid biopsies can here be used as a complement to routine diagnostics.

Analysis of circulating tumor-DNA has been demonstrated to provide clinical value for prognostication, monitoring of treatment efficacy and at an early stage detect relapse. Here, two different digital sequencing strategies have evolved. In the first, medium to large panels are used to target recurrent mutations in a specific tumor type or pan-cancer manner (Phallen et al., 2017). Once developed, these panels are generic and can be used to assess all liquid biopsies. The drawback with this strategy is that the panel needs to target enough sequences to detect any mutation, similarly as in screening. Another concern is that the cost increases with panel size as well. In the second strategy, a tumor tissue sample collected at the time of diagnosis is sequenced with either a panel, whole exome or whole genome setup to identify tumor-specific mutations. To filter out false positives, including germline variants, a non-tumor reference sample collected from the patient, such as DNA extracted from leukocytes in blood, is commonly used. Next, several

identified mutations are selected to design a tumor-specific panel that is used to analyze liquid biopsies from each patient, *i.e.*, personalized panels (McDonald et al., 2019). This strategy requires more efforts before the first liquid biopsy analysis can be performed. However, the sensitivity and specificity are usually superior compared with generic panels, since more target mutations are assessed (Johansson et al., 2019) and the risk of identifying false positives is minimal since no additional targets than confirmed mutations are analyzed (McDonald et al., 2019). Compared to generic panels, personalized panels normally become more attractive when the number of liquid biopsies increases per patient.

Treatment prediction and detection of *de novo* treatment resistance mutations require digital sequencing approaches with generic panels covering mutations linked to available treatment alternatives. On one extreme, such a panel may be wide, including in principle all genes that are currently linked to any treatments when mutated. On the other hand, for some well-defined applications, individual genes or even nucleotide positions may be sufficient to guide clinical decisions since the available treatment options are directly based on mutation status, such as *BRAF*, *EGFR* and *ALK* mutations (García-Pardo et al., 2022).

The use of sequencing in cancer management, including liquid biopsies, is well-established. However, it is less apparent when digital sequencing is required to detect mutations. In most liquid biopsy-based applications there is a need of digital sequencing, enabling detection of variant allele frequencies <0.1%. The amount of circulating tumor-DNA in blood plasma is correlated to tumor burden and stage, where advanced malignancies display high levels of circulating tumor-DNA (Abbosh et al., 2017; Choudhury et al., 2018; Diehl et al., 2005; Newman et al., 2014; Strijker et al., 2020). However, the variation between tumor types and patients is large (Bettegowda et al., 2014). Applications such as screening, minimal residual disease detection and early relapse detection, where the clinical aim normally is to determine whether the patient has cancer as early as possible, are facilitated by the superior sensitivity and specificity of digital sequencing. In other applications, where the patients have confirmed ongoing tumor disease, like diagnostics, prognostication, monitoring treatment efficacy and detection of *de novo* treatment resistance, the choice of method is not as obvious. If the relevant mutation for clinical decision is >1% they may be detected with conventional sequencing approaches. These approaches are often more cost-efficient and enable larger target panels to be analyzed, even whole genomes, that may be used to identify new treatment options. However, as discussed, the amount as well as the relative fraction of tumor-derived DNA is overall lower in liquid biopsies than in tissue biopsies. Therefore, there is substantial risk that relevant mutations are missed despite the fact that the patient has an advanced disease. Hence, the optimal method choice needs to be carefully evaluated for different tumor types and applications. Furthermore, in applications like monitoring of treatment efficacy where the amount of circulating tumor-DNA may vary between high and undetectable levels over time, digital sequencing may be favored. Potentially, when patients are monitored over time digital sequencing may be applied to assess treatment efficacy and for early detection of minimal residual disease and relapse, while standard sequencing targeting large parts of the genome can be used to identify new druggable targets during disease progression. Here, data from the digital sequencing will provide information about the expected amount of circulating tumor-DNA in the sample and hence indicate if the wider but less sensitive standard sequencing will be successful.

2. Conclusions and future perspective

Digital sequencing using UMIs is advancing numerous basic research and clinical applications where standard sequencing fails by enabling sensitive and specific DNA analysis. Successful strategies require careful design and use of UMIs in experimental protocols to allow close to error-free sequencing. We expect more innovative technologies to be established utilizing UMIs that are suitable for current and new application areas, especially for analytes beyond DNA. Digital sequencing requires high sequencing capacity, partly limiting its current use in applications that require large target DNA panels for assessment, such as whole exome and whole genome analysis. To overcome this, new sequencing platforms with higher capacity and lower cost need to be developed. In cancer management, there are several emerging application areas using digital sequencing. However, no single approach is optimal for all applications nor all tumor entities since their technical and clinical requirements are highly variable. We expect that the number of applications where digital sequencing will be implemented in clinical routine will increase substantially in the coming years. To achieve this, true clinical value needs to be demonstrated and that the whole process from sampling, via analysis using digital sequencing, to final clinical data interpretation is aligned with clinical settings, including standardization of the complete workflow.

Declaration of interest

A.S. declares stock ownership and is a board member in Tulebovaasta, Iscaff Pharma and SiMSen Diagnostics. The work has been performed with support from Adlerbertska Foundations; Anna-Lisa och Bror Björnssons Stiftelse; Assar Gabrielssons Foundation; Johan Jansson Foundation for Cancer Research; Magn. Bergvalls Stiftelse; Region Västra Götaland; Swedish Cancer Society [22–2080]; Swedish Childhood Cancer Foundation [2022–0030]; Swedish Research Council [2021–01008]; the Swedish state under the agreement between the Swedish government and the county councils, the ALF-agreement [965065]; Sweden's Innovation Agency [2020–04141]; the Sjöberg Foundation; and the Wilhelm and Martina Lundgren Foundation for Scientific Research.

CRediT authorship contribution statement

Daniel Andersson: Funding acquisition, Investigation, Visualization, Writing – original draft, Writing – review & editing. **Firaol Tamiru Kebede:** Funding acquisition, Investigation, Visualization, Writing – review & editing. **Mandy Escobar:** Funding acquisition, Investigation, Visualization, Writing – review & editing. **Tobias Österlund:** Formal analysis, Funding acquisition, Investigation, Visualization, Writing – original draft, Writing – review & editing. **Anders Ståhlberg:** Conceptualization, Funding acquisition, Investigation, Supervision, Visualization, Writing – original draft, Writing – review & editing.

References

Abascal, F., Harvey, L.M.R., Mitchell, E., Lawson, A.R.J., Lensing, S.V., Ellis, P., et al., 2021. Somatic mutation landscapes at single-molecule resolution. Nature 593 (7859), 405–410. https://doi.org/10.1038/s41586-021-03477-4.

Abbosh, C., Birkbak, N.J., Wilson, G.A., Jamal-Hanjani, M., Constantin, T., Salari, R., et al., 2017. Phylogenetic ctDNA analysis depicts early-stage lung cancer evolution. Nature 545 (7655), 446–451. https://doi.org/10.1038/nature22364.

Andersson, D., Kubista, M., Ståhlberg, A., 2020. Liquid biopsy analysis in cancer diagnostics. Mol. Aspect. Med. 72, 100839 https://doi.org/10.1016/j.mam.2019.100839.

Andrews, T.D., Jeelall, Y., Talaulikar, D., Goodnow, C.C., Field, M.A., 2016. DeepSNVMiner: a sequence analysis tool to detect emergent, rare mutations in subsets of cell populations. PeerJ 4, e2074. https://doi.org/10.7717/peerj.2074.

Arrigoni, L., Ferrari, F., Weller, J., Bella, C., Bönisch, U., Manke, T., 2020. AutoRELACS: automated generation and analysis of ultra-parallel ChIP-seq. Sci. Rep. 10 (1), 12400 https://doi.org/10.1038/s41598-020-69443-8.

Aydemir, O., Janko, M., Hathaway, N.J., Verity, R., Mwandagalirwa, M.K., Tshefu, A.K., et al., 2018. Drug-resistance and population structure of plasmodium falciparum across the democratic republic of Congo using high-throughput molecular inversion probes. J. Infect. Dis. 218 (6), 946–955. https://doi.org/10.1093/infdis/jiy223.

Bae, J.H., Liu, R., Roberts, E., Nguyen, E., Tabrizi, S., Rhoades, J., et al., 2023. Single duplex DNA sequencing with CODEC detects mutations with high sensitivity. Nat. Genet. 55 (5), 871–879. https://doi.org/10.1038/s41588-023-01376-0.

Barnell, E.K., Fisk, B., Skidmore, Z.L., Cotto, K.C., Basu, A., Anand, A., et al., 2022. Personalized ctDNA micro-panels can monitor and predict clinical outcomes for patients with triple-negative breast cancer. Sci. Rep. 12 (1), 17732 https://doi.org/10.1038/s41598-022-20928-8.

Baysoy, A., Bai, Z., Satija, R., Fan, R., 2023. The technological landscape and applications of single-cell multi-omics. Nat. Rev. Mol. Cell Biol. 24 (10), 695–713. https://doi.org/10.1038/s41580-023-00615-w.

Bettegowda, C., Sausen, M., Leary, R.J., Kinde, I., Wang, Y., Agrawal, N., et al., 2014. Detection of circulating tumor DNA in early- and late-stage human malignancies. Sci. Transl. Med. 6 (224), 224ra224 https://doi.org/10.1126/scitranslmed.3007094.

Biddy, B.A., Kong, W., Kamimoto, K., Guo, C., Waye, S.E., Sun, T., et al., 2018. Single-cell mapping of lineage and identity in direct reprogramming. Nature 564 (7735), 219–224. https://doi.org/10.1038/s41586-018-0744-4.

Bolotin, D.A., Poslavsky, S., Mitrophanov, I., Shugay, M., Mamedov, I.Z., Putintseva, E. V., et al., 2015. MiXCR: software for comprehensive adaptive immunity profiling. Nat. Methods 12 (5), 380–381. https://doi.org/10.1038/nmeth.3364.

Bronkhorst, A.J., Ungerer, V., Holdenrieder, S., 2019. The emerging role of cell-free DNA as a molecular marker for cancer management. Biomol Detect Quantif 17, 100087. https://doi.org/10.1016/j.bdq.2019.100087.

Carlson, K.D., Sudmant, P.H., Press, M.O., Eichler, E.E., Shendure, J., Queitsch, C., 2015. MIPSTR: a method for multiplex genotyping of germline and somatic STR variation across many individuals. Genome Res. 25 (5), 750–761. https://doi.org/10.1101/gr.182212.114.

Cescon, D.W., Bratman, S.V., Chan, S.M., Siu, L.L., 2020. Circulating tumor DNA and liquid biopsy in oncology. Nat. Can. (Ott.) 1 (3), 276–290. https://doi.org/10.1038/s43018-020-0043-5.

Chabon, J.J., Hamilton, E.G., Kurtz, D.M., Esfahani, M.S., Moding, E.J., Stehr, H., et al., 2020. Integrating genomic features for non-invasive early lung cancer detection. Nature 580 (7802), 245–251. https://doi.org/10.1038/s41586-020-2140-0.

Chen, L., Liu, P., Evans Jr., T.C., Ettwiller, L.M., 2017. DNA damage is a pervasive cause of sequencing errors, directly confounding variant identification. Science 355 (6326), 752–756. https://doi.org/10.1126/science.aai8690.

Choudhury, A.D., Werner, L., Francini, E., Wei, X.X., Ha, G., Freeman, S.S., et al., 2018. Tumor fraction in cell-free DNA as a biomarker in prostate cancer. JCI Insight 3 (21). https://doi.org/10.1172/jci.insight.122109.

Chung, J., Lee, K.W., Lee, C., Shin, S.H., Kyung, S., Jeon, H.J., et al., 2019. Performance evaluation of commercial library construction kits for PCR-based targeted sequencing using a unique molecular identifier. BMC Genom. 20 (1), 216. https://doi.org/10.1186/s12864-019-5583-7.

Cohen, J.D., Douville, C., Dudley, J.C., Mog, B.J., Popoli, M., Ptak, J., et al., 2021. Detection of low-frequency DNA variants by targeted sequencing of the Watson and Crick strands. Nat. Biotechnol. 39 (10), 1220–1227. https://doi.org/10.1038/s41587-021-00900-z.

Cohen, J.D., Li, L., Wang, Y., Thoburn, C., Afsari, B., Danilova, L., et al., 2018. Detection and localization of surgically resectable cancers with a multi-analyte blood test. Science 359 (6378), 926–930. https://doi.org/10.1126/science.aar3247.

Costello, M., Pugh, T.J., Fennell, T.J., Stewart, C., Lichtenstein, L., Meldrim, J.C., et al., 2013. Discovery and characterization of artifactual mutations in deep coverage targeted capture sequencing data due to oxidative DNA damage during sample preparation. Nucleic Acids Res. 41 (6), e67. https://doi.org/10.1093/nar/gks1443.

Derrien, J., Gastineau, S., Frigout, A., Giordano, N., Cherkaoui, M., Gaborit, V., et al., 2023. Acquired resistance to a GPRC5D-directed T-cell engager in multiple myeloma is mediated by genetic or epigenetic target inactivation. Nat. Can. (Ott.) 4 (11), 1536–1543. https://doi.org/10.1038/s43018-023-00625-9.

Diehl, F., Li, M., Dressman, D., He, Y., Shen, D., Szabo, S., et al., 2005. Detection and quantification of mutations in the plasma of patients with colorectal tumors. Proc. Natl. Acad. Sci. U. S. A. 102 (45), 16368–16373. https://doi.org/10.1073/pnas.0507904102.

Do, H., Dobrovic, A., 2015. Sequence artifacts in DNA from formalin-fixed tissues: causes and strategies for minimization. Clin. Chem. 61 (1), 64–71. https://doi.org/10.1373/clinchem.2014.223040.

Dodge, A.E., LeBlanc, D.P.M., Zhou, G., Williams, A., Meier, M.J., Van, P., et al., 2023. Duplex sequencing provides detailed characterization of mutation frequencies and spectra in the bone marrow of MutaMouse males exposed to procarbazine hydrochloride. Arch. Toxicol. 97 (8), 2245–2259. https://doi.org/10.1007/s00204-023-03527-y.

Egyud, M., Sridhar, P., Devaiah, A., Yamada, E., Saunders, S., Ståhlberg, A., et al., 2019. Plasma circulating tumor DNA as a potential tool for disease monitoring in head and neck cancer. Head Neck 41 (5), 1351–1358. https://doi.org/10.1002/hed.25563.

Fields, B., Moeskjaer, S., Friman, V.P., Andersen, S.U., Young, J.P.W., 2021. MAUI-seq: metabarcoding using amplicons with unique molecular identifiers to improve error correction. Mol Ecol Resour 21 (3), 703–720. https://doi.org/10.1111/1755-0998.13294.

Filges, S., Mouhanna, P., Ståhlberg, A., 2021. Digital quantification of chemical oligonucleotide synthesis errors. Clin. Chem. 67 (10), 1384–1394. https://doi.org/10.1093/clinchem/hvab136.

Filges, S., Yamada, E., Ståhlberg, A., Godfrey, T.E., 2019. Impact of polymerase fidelity on background error rates in next-generation sequencing with unique molecular identifiers/barcodes. Sci. Rep. 9 (1), 3503. https://doi.org/10.1038/s41598-019-39762-6.

Flajolet, P., Gardy, D., Thimonier, L., 1992. Birthday paradox, coupon collectors, caching algorithms and self-organizing search. Discrete Appl. Math. 39 (3), 207–229. https://doi.org/10.1016/0166-218x(92)90177-c.

Fox, E.J., Reid-Bayliss, K.S., Emond, M.J., Loeb, L.A., 2014. Accuracy of next generation sequencing platforms. Next Gener Seq Appl 1. https://doi.org/10.4172/jngsa.1000106.

Fu, C., Marczyk, M., Samuels, M., Trevarton, A.J., Qu, J., Lau, R., et al., 2021. Targeted RNAseq assay incorporating unique molecular identifiers for improved quantification of gene expression signatures and transcribed mutation fraction in fixed tumor samples. BMC Cancer 21 (1), 114. https://doi.org/10.1186/s12885-021-07814-8.

Fu, Y., Wu, P.H., Beane, T., Zamore, P.D., Weng, Z., 2018. Elimination of PCR duplicates in RNA-seq and small RNA-seq using unique molecular identifiers. BMC Genom. 19 (1), 531. https://doi.org/10.1186/s12864-018-4933-1.

Gao, M., Ling, M., Tang, X., Wang, S., Xiao, X., Qiao, Y., et al., 2021. Comparison of high-throughput single-cell RNA sequencing data processing pipelines. Briefings Bioinf 22 (3), bbaa116. https://doi.org/10.1093/bib/bbaa116.

García-Pardo, M., Makarem, M., Li, J.J.N., Kelly, D., Leighl, N.B., 2022. Integrating circulating-free DNA (cfDNA) analysis into clinical practice: opportunities and challenges. Br. J. Cancer 127 (4), 592–602. https://doi.org/10.1038/s41416-022-01776-9.

George, S., Suwondo, P., Akorli, J., Otchere, J., Harrison, L.M., Bilguvar, K., et al., 2022. Application of multiplex amplicon deep-sequencing (MAD-seq) to screen for putative drug resistance markers in the Necator americanus isotype-1 β-tubulin gene. Sci. Rep. 12 (1), 11459 https://doi.org/10.1038/s41598-022-15718-1.

Ghiam, B.K., Xu, L., Berry, J.L., 2019. Aqueous humor markers in retinoblastoma, a review. Transl Vis Sci Technol 8 (2), 13. https://doi.org/10.1167/tvst.8.2.13.

Gregory, M.T., Bertout, J.A., Ericson, N.G., Taylor, S.D., Mukherjee, R., Robins, H.S., et al., 2016. Targeted single molecule mutation detection with massively parallel sequencing. Nucleic Acids Res. 44 (3), e22. https://doi.org/10.1093/nar/gkv915.

Gröbner, S.N., Worst, B.C., Weischenfeldt, J., Buchhalter, I., Kleinheinz, K., Rudneva, V. A., et al., 2018. The landscape of genomic alterations across childhood cancers. Nature 555 (7696), 321–327. https://doi.org/10.1038/nature25480.

Gydush, G., Nguyen, E., Bae, J.H., Blewett, T., Rhoades, J., Reed, S.C., et al., 2022. Massively parallel enrichment of low-frequency alleles enables duplex sequencing at low depth. Nat. Biomed. Eng. 6 (3), 257–266. https://doi.org/10.1038/s41551-022-00855-9.

Hagemann-Jensen, M., Ziegenhain, C., Chen, P., Ramsköld, D., Hendriks, G.J., Larsson, A.J.M., et al., 2020. Single-cell RNA counting at allele and isoform resolution using Smart-seq3. Nat. Biotechnol. 38 (6), 708–714. https://doi.org/10.1038/s41587-020-0497-0.

Heitzer, E., Auinger, L., Speicher, M.R., 2020. Cell-free DNA and apoptosis: how dead cells inform about the living. Trends Mol. Med. 26 (5), 519–528. https://doi.org/10.1016/j.molmed.2020.01.012.

Heumos, L., Schaar, A.C., Lance, C., Litinetskaya, A., Drost, F., Zappia, L., et al., 2023. Best practices for single-cell analysis across modalities. Nat. Rev. Genet. 24 (8), 550–572. https://doi.org/10.1038/s41576-023-00586-w.

Hiatt, J.B., Patwardhan, R.P., Turner, E.H., Lee, C., Shendure, J., 2010. Parallel, tag-directed assembly of locally derived short sequence reads. Nat. Methods 7 (2), 119–122. https://doi.org/10.1038/nmeth.1416.

Hiatt, J.B., Pritchard, C.C., Salipante, S.J., O'Roak, B.J., Shendure, J., 2013. Single molecule molecular inversion probes for targeted, high-accuracy detection of low-frequency variation. Genome Res. 23 (5), 843–854. https://doi.org/10.1101/gr.147686.112.

Hilke, F.J., Muyas, F., Admard, J., Kootz, B., Nann, D., Welz, S., et al., 2020. Dynamics of cell-free tumour DNA correlate with treatment response of head and neck cancer patients receiving radiochemotherapy. Radiother. Oncol. 151, 182–189. https://doi.org/10.1016/j.radonc.2020.07.027.

Hoang, M.L., Kinde, I., Tomasetti, C., McMahon, K.W., Rosenquist, T.A., Grollman, A.P., et al., 2016. Genome-wide quantification of rare somatic mutations in normal human tissues using massively parallel sequencing. Proc. Natl. Acad. Sci. U. S. A. 113 (35), 9846–9851. https://doi.org/10.1073/pnas.1607794113.

Hoekstra, J.G., Hipp, M.J., Montine, T.J., Kennedy, S.R., 2016. Mitochondrial DNA mutations increase in early stage Alzheimer disease and are inconsistent with oxidative damage. Ann. Neurol. 80 (2), 301–306. https://doi.org/10.1002/ana.24709.

Hong, J., Gresham, D., 2017. Incorporation of unique molecular identifiers in TruSeq adapters improves the accuracy of quantitative sequencing. Biotechniques 63 (5), 221–226. https://doi.org/10.2144/000114608.

Hu, L., Su, L., Cheng, H., Mo, C., Ouyang, T., Li, J., et al., 2021. Single-cell RNA sequencing reveals the cellular origin and evolution of breast cancer in BRCA1 mutation carriers. Cancer Res. 81 (10), 2600–2611. https://doi.org/10.1158/0008-5472.CAN-20-2123.

Hug, H., Schuler, R., 2003. Measurement of the number of molecules of a single mRNA species in a complex mRNA preparation. J. Theor. Biol. 221 (4), 615–624. https://doi.org/10.1006/jtbi.2003.3211.

Hwang, B., Lee, J.H., Bang, D., 2018. Single-cell RNA sequencing technologies and bioinformatics pipelines. Exp. Mol. Med. 50 (8), 1–14. https://doi.org/10.1038/s12276-018-0071-8.

Ivanov, M., Baranova, A., Butler, T., Spellman, P., Mileyko, V., 2015. Non-random fragmentation patterns in circulating cell-free DNA reflect epigenetic regulation. BMC Genom. 16 (Suppl. 13), S1. https://doi.org/10.1186/1471-2164-16-S13-S1. Suppl 13.

Jäger, A.C., Alvarez, M.L., Davis, C.P., Guzmán, E., Han, Y., Way, L., et al., 2017. Developmental validation of the MiSeq FGx forensic genomics system for targeted next generation sequencing in forensic DNA casework and database laboratories. Forensic Sci Int Genet 28, 52–70. https://doi.org/10.1016/j.fsigen.2017.01.011.

Johansson, G., Andersson, D., Filges, S., Li, J., Muth, A., Godfrey, T.E., et al., 2019. Considerations and quality controls when analyzing cell-free tumor DNA. Biomol Detect Quantif 17, 100078. https://doi.org/10.1016/j.bdq.2018.12.003.

Johansson, G., Berndsen, M., Lindskog, S., Österlund, T., Fagman, H., Muth, A., et al., 2021. Monitoring circulating tumor DNA during surgical treatment in patients with gastrointestinal stromal tumors. Mol. Cancer Therapeut. 20 (12), 2568–2576. https://doi.org/10.1158/1535-7163.MCT-21-0403.

Johansson, G., Kaltak, M., Rimniceanu, C., Singh, A.K., Lycke, J., Malmeström, C., et al., 2020. Ultrasensitive DNA immune repertoire sequencing using unique molecular identifiers. Clin. Chem. 66 (9), 1228–1237. https://doi.org/10.1093/clinchem/hvaa159.

Johnson, M.S., Venkataram, S., Kryazhimskiy, S., 2023. Best practices in designing, sequencing, and identifying random DNA barcodes. J. Mol. Evol. 91 (3), 263–280. https://doi.org/10.1007/s00239-022-10083-z.

Jovic, D., Liang, X., Zeng, H., Lin, L., Xu, F., Luo, Y., 2022. Single-cell RNA sequencing technologies and applications: a brief overview. Clin. Transl. Med. 12 (3), e694. https://doi.org/10.1002/ctm2.694.

Kargapolova, Y., Levin, M., Lackner, K., Danckwardt, S., 2017. sCLIP-an integrated platform to study RNA-protein interactomes in biomedical research: identification of CSTF2tau in alternative processing of small nuclear RNAs. Nucleic Acids Res. 45 (10), 6074–6086. https://doi.org/10.1093/nar/gkx152.

Karst, S.M., Ziels, R.M., Kirkegaard, R.H., Sørensen, E.A., McDonald, D., Zhu, Q., et al., 2021. High-accuracy long-read amplicon sequences using unique molecular identifiers with Nanopore or PacBio sequencing. Nat. Methods 18 (2), 165–169. https://doi.org/10.1038/s41592-020-01041-y.

Kennedy, S.R., Schmitt, M.W., Fox, E.J., Kohrn, B.F., Salk, J.J., Ahn, E.H., et al., 2014. Detecting ultralow-frequency mutations by duplex sequencing. Nat. Protoc. 9 (11), 2586–2606. https://doi.org/10.1038/nprot.2014.170.

Keys, J.R., Zhou, S., Anderson, J.A., Eron Jr., J.J., Rackoff, L.A., Jabara, C., et al., 2015. Primer ID informs next-generation sequencing platforms and reveals preexisting drug resistance mutations in the HIV-1 reverse transcriptase coding domain. AIDS Res. Hum. Retrovir. 31 (6), 658–668. https://doi.org/10.1089/AID.2014.0031.

Kinde, I., Wu, J., Papadopoulos, N., Kinzler, K.W., Vogelstein, B., 2011. Detection and quantification of rare mutations with massively parallel sequencing. Proc. Natl. Acad. Sci. U. S. A. 108 (23), 9530–9535. https://doi.org/10.1073/pnas.1105422108.

Ko, S.H., Bayat Mokhtari, E., Mudvari, P., Stein, S., Stringham, C.D., Wagner, D., et al., 2021. High-throughput, single-copy sequencing reveals SARS-CoV-2 spike variants coincident with mounting humoral immunity during acute COVID-19. PLoS Pathog. 17 (4), e1009431 https://doi.org/10.1371/journal.ppat.1009431.

Koh, C.W.Q., Goh, Y.T., Goh, W.S.S., 2019. Atlas of quantitative single-base-resolution N^6-methyl-adenine methylomes. Nat. Commun. 10 (1), 5636. https://doi.org/10.1038/s41467-019-13561-z.

König, J., Zarnack, K., Rot, G., Curk, T., Kayikci, M., Zupan, B., et al., 2010. iCLIP reveals the function of hnRNP particles in splicing at individual nucleotide resolution. Nat. Struct. Mol. Biol. 17 (7), 909–915. https://doi.org/10.1038/nsmb.1838.

Kueng, N., Arcioni, S., Sandberg, F., Kuhn, C., Banz, V., Largiadèr, C.R., et al., 2023. Comparison of methods for donor-derived cell-free DNA quantification in plasma and urine from solid organ transplant recipients. Front. Genet. 14, 1089830 https://doi.org/10.3389/fgene.2023.1089830.

Kukita, Y., Matoba, R., Uchida, J., Hamakawa, T., Doki, Y., Imamura, F., et al., 2015. High-fidelity target sequencing of individual molecules identified using barcode sequences: de novo detection and absolute quantitation of mutations in plasma cell-free DNA from cancer patients. DNA Res. 22 (4), 269–277. https://doi.org/10.1093/dnares/dsv010.

Kumar, B., Elsässer, S.J., 2019. Quantitative multiplexed ChIP reveals global alterations that shape promoter bivalency in ground state embryonic stem cells. Cell Rep. 28 (12), 3274–3284. https://doi.org/10.1016/j.celrep.2019.08.046 e3275.

Levin, I., Štrajbl, M., Fastman, Y., Baran, D., Twito, S., Mioduser, J., et al., 2023. Accurate profiling of full-length Fv in highly homologous antibody libraries using UMI tagged short reads. Nucleic Acids Res. 51 (11), e61. https://doi.org/10.1093/nar/gkad235.

Liu, X., Liu, L., Ji, Y., Li, C., Wei, T., Yang, X., et al., 2019. Enrichment of short mutant cell-free DNA fragments enhanced detection of pancreatic cancer. EBioMedicine 41, 345–356. https://doi.org/10.1016/j.ebiom.2019.02.010.

Lo, Y.M., Zhang, J., Leung, T.N., Lau, T.K., Chang, A.M., Hjelm, N.M., 1999. Rapid clearance of fetal DNA from maternal plasma. Am. J. Hum. Genet. 64 (1), 218–224. https://doi.org/10.1086/302205.

Lu, C., Zhu, H., Zhao, D., Zhang, J., Yang, K., Lv, Y., et al., 2022. Oral-Gut microbiome analysis in patients with metabolic-associated fatty liver disease having different tongue image feature. Front. Cell. Infect. Microbiol. 12, 787143 https://doi.org/10.3389/fcimb.2022.787143.

Luecken, M.D., Theis, F.J., 2019. Current best practices in single-cell RNA-seq analysis: a tutorial. Mol. Syst. Biol. 15 (6), e8746 https://doi.org/10.15252/msb.20188746.

Mansukhani, S., Barber, L.J., Kleftogiannis, D., Moorcraft, S.Y., Davidson, M., Woolston, A., et al., 2018. Ultra-sensitive mutation detection and genome-wide DNA copy number reconstruction by error-corrected circulating tumor DNA sequencing. Clin. Chem. 64 (11), 1626–1635. https://doi.org/10.1373/clinchem.2018.289629.

Mármol-Sánchez, E., Fromm, B., Oskolkov, N., Pochon, Z., Kalogeropoulos, P., Eriksson, E., et al., 2023. Historical RNA expression profiles from the extinct Tasmanian tiger. Genome Res. 33 (8), 1299–1316. https://doi.org/10.1101/gr.277663.123.

Maslov, A.Y., Makhortov, S., Sun, S., Heid, J., Dong, X., Lee, M., et al., 2022. Single-molecule, quantitative detection of low-abundance somatic mutations by high-throughput sequencing. Sci. Adv. 8 (14), eabm3259 https://doi.org/10.1126/sciadv.abm3259.

Mattox, A.K., Wang, Y., Springer, S., Cohen, J.D., Yegnasubramanian, S., Nelson, W.G., et al., 2017. Bisulfite-converted duplexes for the strand-specific detection and quantification of rare mutations. Proc. Natl. Acad. Sci. U. S. A. 114 (18), 4733–4738. https://doi.org/10.1073/pnas.1701382114.

McDonald, B.R., Contente-Cuomo, T., Sammut, S.J., Odenheimer-Bergman, A., Ernst, B., Perdigones, N., et al., 2019. Personalized circulating tumor DNA analysis to detect residual disease after neoadjuvant therapy in breast cancer. Sci. Transl. Med. 11 (504) https://doi.org/10.1126/scitranslmed.aax7392.

Mead, R., Duku, M., Bhandari, P., Cree, I.A., 2011. Circulating tumour markers can define patients with normal colons, benign polyps, and cancers. Br. J. Cancer 105 (2), 239–245. https://doi.org/10.1038/bjc.2011.230.

Melsted, P., Ntranos, V., Pachter, L., 2019. The barcode, UMI, set format and BUStools. Bioinformatics 35 (21), 4472–4473. https://doi.org/10.1093/bioinformatics/btz279.

Moonsamy, D., Gentle, N., 2022. SASCRiP: a Python workflow for preprocessing UMI count-based scRNA-seq data. F1000Research 11 (190), 190. https://doi.org/10.12688/f1000research.75243.1.

Moss, J., Magenheim, J., Neiman, D., Zemmour, H., Loyfer, N., Korach, A., et al., 2018. Comprehensive human cell-type methylation atlas reveals origins of circulating cell-free DNA in health and disease. Nat. Commun. 9 (1), 5068. https://doi.org/10.1038/s41467-018-07466-6.

Mouliere, F., El Messaoudi, S., Pang, D., Dritschilo, A., Thierry, A.R., 2014. Multi-marker analysis of circulating cell-free DNA toward personalized medicine for colorectal cancer. Mol. Oncol. 8 (5), 927–941. https://doi.org/10.1016/j.molonc.2014.02.005.

Narayan, A., Carriero, N.J., Gettinger, S.N., Kluytenaar, J., Kozak, K.R., Yock, T.I., et al., 2012. Ultrasensitive measurement of hotspot mutations in tumor DNA in blood using error-suppressed multiplexed deep sequencing. Cancer Res. 72 (14), 3492–3498. https://doi.org/10.1158/0008-5472.CAN-11-4037.

Newman, A.M., Bratman, S.V., To, J., Wynne, J.F., Eclov, N.C., Modlin, L.A., et al., 2014. An ultrasensitive method for quantitating circulating tumor DNA with broad patient coverage. Nat. Med. 20 (5), 548–554. https://doi.org/10.1038/nm.3519.

Newman, A.M., Lovejoy, A.F., Klass, D.M., Kurtz, D.M., Chabon, J.J., Scherer, F., et al., 2016. Integrated digital error suppression for improved detection of circulating tumor DNA. Nat. Biotechnol. 34 (5), 547–555. https://doi.org/10.1038/nbt.3520.

Noyes, N.R., Weinroth, M.E., Parker, J.K., Dean, C.J., Lakin, S.M., Raymond, R.A., et al., 2017. Enrichment allows identification of diverse, rare elements in metagenomic resistome-virulome sequencing. Microbiome 5 (1), 142. https://doi.org/10.1186/s40168-017-0361-8.

Nurk, S., Koren, S., Rhie, A., Rautiainen, M., Bzikadze, A.V., Mikheenko, A., et al., 2022. The complete sequence of a human genome. Science 376 (6588), 44–53. https://doi.org/10.1126/science.abj6987.

Orabi, B., Erhan, E., McConeghy, B., Volik, S.V., Le Bihan, S., Bell, R., et al., 2019. Alignment-free clustering of UMI tagged DNA molecules. Bioinformatics 35 (11), 1829–1836. https://doi.org/10.1093/bioinformatics/bty888.

Österlund, T., Filges, S., Johansson, G., Ståhlberg, A., 2022. UMIErrorCorrect and UMIAnalyzer: software for consensus read generation, error correction, and visualization using unique molecular identifiers. Clin. Chem. 68 (11), 1425–1435. https://doi.org/10.1093/clinchem/hvac136.

Pandey, S., Shen, K., Lee, S.H., Shen, Y.A., Wang, Y., Otero-García, M., et al., 2022. Disease-associated oligodendrocyte responses across neurodegenerative diseases. Cell Rep. 40 (8), 111189 https://doi.org/10.1016/j.celrep.2022.111189.

Patkar, N., Kakirde, C., Shaikh, A.F., Salve, R., Bhanshe, P., Chatterjee, G., et al., 2021. Clinical impact of panel-based error-corrected next generation sequencing versus flow cytometry to detect measurable residual disease (MRD) in acute myeloid leukemia (AML). Leukemia 35 (5), 1392–1404. https://doi.org/10.1038/s41375-021-01131-6.

Pel, J., Leung, A., Choi, W.W.Y., Despotovic, M., Ung, W.L., Shibahara, G., et al., 2018. Rapid and highly-specific generation of targeted DNA sequencing libraries enabled by linking capture probes with universal primers. PLoS One 13 (12), e0208283. https://doi.org/10.1371/journal.pone.0208283.

Peng, X., Dorman, K.S., 2023. Accurate estimation of molecular counts from amplicon sequence data with unique molecular identifiers. Bioinformatics 39 (1), btad002. https://doi.org/10.1093/bioinformatics/btad002.

Phallen, J., Sausen, M., Adleff, V., Leal, A., Hruban, C., White, J., et al., 2017. Direct detection of early-stage cancers using circulating tumor DNA. Sci. Transl. Med. 9 (403) https://doi.org/10.1126/scitranslmed.aan2415.

Piovesan, A., Pelleri, M.C., Antonaros, F., Strippoli, P., Caracausi, M., Vitale, L., 2019. On the length, weight and GC content of the human genome. BMC Res. Notes 12 (1), 106. https://doi.org/10.1186/s13104-019-4137-z.

Qiu, B., Guo, W., Zhang, F., Lv, F., Ji, Y., Peng, Y., et al., 2021. Dynamic recurrence risk and adjuvant chemotherapy benefit prediction by ctDNA in resected NSCLC. Nat. Commun. 12 (1), 6770. https://doi.org/10.1038/s41467-021-27022-z.

Razavi, P., Li, B.T., Brown, D.N., Jung, B., Hubbell, E., Shen, R., et al., 2019. High-intensity sequencing reveals the sources of plasma circulating cell-free DNA variants. Nat. Med. 25 (12), 1928–1937. https://doi.org/10.1038/s41591-019-0652-7.

Ryoo, S.B., Heo, S., Lim, Y., Lee, W., Cho, S.H., Ahn, J., et al., 2023. Personalised circulating tumour DNA assay with large-scale mutation coverage for sensitive minimal residual disease detection in colorectal cancer. Br. J. Cancer 129 (2), 374–381. https://doi.org/10.1038/s41416-023-02300-3.

Sagitov, S., Ståhlberg, A., 2023. Counting unique molecular identifiers in sequencing using a multi-type branching process with immigration. J. Theor. Biol. 558, 111365 https://doi.org/10.1016/j.jtbi.2022.111365.

Salk, J.J., Loubet-Senear, K., Maritschnegg, E., Valentine, C.C., Williams, L.N., Higgins, J. E., et al., 2019. Ultra-sensitive TP53 sequencing for cancer detection reveals progressive clonal selection in normal tissue over a century of human lifespan. Cell Rep. 28 (1), 132–144. https://doi.org/10.1016/j.celrep.2019.05.109 e133.

Salk, J.J., Schmitt, M.W., Loeb, L.A., 2018. Enhancing the accuracy of next-generation sequencing for detecting rare and subclonal mutations. Nat. Rev. Genet. 19 (5), 269–285. https://doi.org/10.1038/nrg.2017.117.

Sater, V., Viailly, P.J., Lecroq, T., Prieur-Gaston, É., Bohers, É., Viennot, M., et al., 2020. UMI-VarCal: a new UMI-based variant caller that efficiently improves low-frequency variant detection in paired-end sequencing NGS libraries. Bioinformatics 36 (9), 2718–2724. https://doi.org/10.1093/bioinformatics/btaa05.

Schmitt, M.W., Fox, E.J., Prindle, M.J., Reid-Bayliss, K.S., True, L.D., Radich, J.P., et al., 2015. Sequencing small genomic targets with high efficiency and extreme accuracy. Nat. Methods 12 (5), 423–425. https://doi.org/10.1038/nmeth.3351.

Schmitt, M.W., Kennedy, S.R., Salk, J.J., Fox, E.J., Hiatt, J.B., Loeb, L.A., 2012. Detection of ultra-rare mutations by next-generation sequencing. Proc. Natl. Acad. Sci. U. S. A. 109 (36), 14508–14513. https://doi.org/10.1073/pnas.1208715109.

Shugay, M., Britanova, O.V., Merzlyak, E.M., Turchaninova, M.A., Mamedov, I.Z., Tuganbaev, T.R., et al., 2014. Towards error-free profiling of immune repertoires. Nat. Methods 11 (6), 653–655. https://doi.org/10.1038/nmeth.2960.

Shugay, M., Zaretsky, A.R., Shagin, D.A., Shagina, I.A., Volchenkov, I.A., Shelenkov, A. A., et al., 2017. MAGERI: computational pipeline for molecular-barcoded targeted resequencing. PLoS Comput. Biol. 13 (5), e1005480 https://doi.org/10.1371/journal.pcbi.1005480.

Sidstedt, M., Rådstrom, P., Hedman, J., 2020. PCR inhibition in qPCR, dPCR and MPS-mechanisms and solutions. Anal. Bioanal. Chem. 412 (9), 2009–2023. https://doi.org/10.1007/s00216-020-02490-2.

Smith, T., Heger, A., Sudbery, I., 2017. UMI-tools: modeling sequencing errors in Unique Molecular Identifiers to improve quantification accuracy. Genome Res. 27 (3), 491–499. https://doi.org/10.1101/gr.209601.116.

Smyczynska, U., Bartlomiejczyk, M.A., Stanczak, M.M., Sztromwasser, P., Wesolowska, A., Barbarska, O., et al., 2020. Impact of processing method on donated human breast milk microRNA content. PLoS One 15 (7), e0236126. https://doi.org/10.1371/journal.pone.0236126.

Snyder, M.W., Kircher, M., Hill, A.J., Daza, R.M., Shendure, J., 2016. Cell-free DNA comprises an in vivo nucleosome footprint that informs its tissues-of-origin. Cell 164 (1–2), 57–68. https://doi.org/10.1016/j.cell.2015.11.050.

Stackpole, M.L., Zeng, W., Li, S., Liu, C.C., Zhou, Y., He, S., et al., 2022. Cost-effective methylome sequencing of cell-free DNA for accurately detecting and locating cancer. Nat. Commun. 13 (1), 5566. https://doi.org/10.1038/s41467-022-32995-6.

Ståhlberg, A., Krzyzanowski, P.M., Egyud, M., Filges, S., Stein, L., Godfrey, T.E., 2017. Simple multiplexed PCR-based barcoding of DNA for ultrasensitive mutation detection by next-generation sequencing. Nat. Protoc. 12 (4), 664–682. https://doi.org/10.1038/nprot.2017.006.

Ståhlberg, A., Krzyzanowski, P.M., Jackson, J.B., Egyud, M., Stein, L., Godfrey, T.E., 2016. Simple, multiplexed, PCR-based barcoding of DNA enables sensitive mutation detection in liquid biopsies using sequencing. Nucleic Acids Res. 44 (11), e105. https://doi.org/10.1093/nar/gkw224.

Stankunaite, R., George, S.L., Gallagher, L., Jamal, S., Shaikh, R., Yuan, L., et al., 2022. Circulating tumour DNA sequencing to determine therapeutic response and identify tumour heterogeneity in patients with paediatric solid tumours. Eur. J. Cancer 162, 209–220. https://doi.org/10.1016/j.ejca.2021.09.042.

Stead, L.F., Sutton, K.M., Taylor, G.R., Quirke, P., Rabbitts, P., 2013. Accurately identifying low-allelic fraction variants in single samples with next-generation sequencing: applications in tumor subclone resolution. Hum. Mutat. 34 (10), 1432–1438. https://doi.org/10.1002/humu.22365.

Stetson, D., Ahmed, A., Xu, X., Nuttall, B.R.B., Lubinski, T.J., Johnson, J.H., et al., 2019. Orthogonal comparison of four plasma NGS tests with tumor suggests technical factors are a major source of assay discordance. JCO Precis Oncol 3, 1–9. https://doi.org/10.1200/PO.18.00191.

Strijker, M., Soer, E.C., de Pastena, M., Creemers, A., Balduzzi, A., Beagan, J.J., et al., 2020. Circulating tumor DNA quantity is related to tumor volume and both predict survival in metastatic pancreatic ductal adenocarcinoma. Int. J. Cancer 146 (5), 1445–1456. https://doi.org/10.1002/ijc.32586.

Suppan, C., Graf, R., Jahn, S., Zhou, Q., Klocker, E.V., Bartsch, R., et al., 2022. Sensitive and robust liquid biopsy-based detection of PIK3CA mutations in hormone-receptor-positive metastatic breast cancer patients. Br. J. Cancer 126 (3), 456–463. https://doi.org/10.1038/s41416-021-01601-9.

Tam, J.C.W., Chan, Y.M., Tsang, S.Y., Yau, C.I., Yeung, S.Y., Au, K.K., et al., 2020. Noninvasive prenatal paternity testing by means of SNP-based targeted sequencing. Prenat. Diagn. 40 (4), 497–506. https://doi.org/10.1002/pd.5595.

Taylor, B.S., Barretina, J., Maki, R.G., Antonescu, C.R., Singer, S., Ladanyi, M., 2011. Advances in sarcoma genomics and new therapeutic targets. Nat. Rev. Cancer 11 (8), 541–557. https://doi.org/10.1038/nrc3087.

Tie, J., Wang, Y., Tomasetti, C., Li, L., Springer, S., Kinde, I., et al., 2016. Circulating tumor DNA analysis detects minimal residual disease and predicts recurrence in patients with stage II colon cancer. Sci. Transl. Med. 8 (346), 346ra392 https://doi.org/10.1126/scitranslmed.aaf6219.

To, E.W.H., Chan, K.C.A., Leung, S.F., Chan, L.Y.S., To, K.F., Chan, A.T.C., et al., 2003. Rapid clearance of plasma Epstein-Barr virus DNA after surgical treatment of nasopharyngeal carcinoma. Clin. Cancer Res. 9 (9), 3254–3259.

Tsagiopoulou, M., Maniou, M.C., Pechlivanis, N., Togkousidis, A., Kotrová, M., Hutzenlaub, T., et al., 2021. UMIc: a preprocessing method for UMI deduplication and reads correction. Front. Genet. 12, 660366 https://doi.org/10.3389/fgene.2021.660366.

Valentine 3rd, C.C., Young, R.R., Fielden, M.R., Kulkarni, R., Williams, L.N., Li, T., et al., 2020. Direct quantification of in vivo mutagenesis and carcinogenesis using duplex sequencing. Proc. Natl. Acad. Sci. U. S. A. 117 (52), 33414–33425. https://doi.org/10.1073/pnas.2013724117.

Van Nostrand, E.L., Pratt, G.A., Shishkin, A.A., Gelboin-Burkhart, C., Fang, M.Y., Sundararaman, B., et al., 2016. Robust transcriptome-wide discovery of RNA-binding protein binding sites with enhanced CLIP (eCLIP). Nat. Methods 13 (6), 508–514. https://doi.org/10.1038/nmeth.3810.

Varghese, A.M., Patel, J., Janjigian, Y.Y., Meng, F., Selcuklu, S.D., Iyer, G., et al., 2021. Noninvasive detection of polyclonal acquired resistance to FGFR inhibition in patients with cholangiocarcinoma harboring FGFR2 alterations. JCO Precis Oncol 5. https://doi.org/10.1200/PO.20.00178.

Varley, K.E., Gertz, J., Bowling, K.M., Parker, S.L., Reddy, T.E., Pauli-Behn, F., et al., 2013. Dynamic DNA methylation across diverse human cell lines and tissues. Genome Res. 23 (3), 555–567. https://doi.org/10.1101/gr.147942.112.

Vogelstein, B., Papadopoulos, N., Velculescu, V.E., Zhou, S., Diaz Jr., L.A., Kinzler, K.W., 2013. Cancer genome landscapes. Science 339 (6127), 1546–1558. https://doi.org/10.1126/science.1235122.

Waldeck, S., Mitschke, J., Wiesemann, S., Rassner, M., Andrieux, G., Deuter, M., et al., 2022. Early assessment of circulating tumor DNA after curative-intent resection predicts tumor recurrence in early-stage and locally advanced non-small-cell lung cancer. Mol. Oncol. 16 (2), 527–537. https://doi.org/10.1002/1878-0261.13116.

Walens, A., Lin, J., Damrauer, J.S., McKinney, B., Lupo, R., Newcomb, R., et al., 2020. Adaptation and selection shape clonal evolution of tumors during residual disease and recurrence. Nat. Commun. 11 (1), 5017. https://doi.org/10.1038/s41467-020-18730-z.

Wang, Q., Wang, X., Tang, P.S., O'leary, G.M., Zhang, M., 2017. Targeted sequencing of both DNA strands barcoded and captured individually by RNA probes to identify genome-wide ultra-rare mutations. Sci. Rep. 7 (1), 3356. https://doi.org/10.1038/s41598-017-03448-8.

Weber, S., Spiegl, B., Perakis, S.O., Ulz, C.M., Abuja, P.M., Kashofer, K., et al., 2020. Technical evaluation of commercial mutation analysis platforms and reference materials for liquid biopsy profiling. Cancers 12 (6). https://doi.org/10.3390/cancers12061588.

Wu, D., Sherwood, A., Fromm, J.R., Winter, S.S., Dunsmore, K.P., Loh, M.L., et al., 2012. High-throughput sequencing detects minimal residual disease in acute T lymphoblastic leukemia. Sci. Transl. Med. 4 (134), 134ra163 https://doi.org/10.1126/scitranslmed.3003656.

Wu, L., Chu, X., Zheng, J., Xiao, C., Zhang, Z., Huang, G., et al., 2019. Targeted capture and sequencing of 1245 SNPs for forensic applications. Forensic Sci Int Genet 42, 227–234. https://doi.org/10.1016/j.fsigen.2019.07.006.

Xia, B., Yan, Y., Baron, M., Wagner, F., Barkley, D., Chiodin, M., et al., 2020. Widespread transcriptional scanning in the testis modulates gene evolution rates. Cell 180 (2), 248–262. https://doi.org/10.1016/j.cell.2019.12.015 e221.

Xu, C., Gu, X., Padmanabhan, R., Wu, Z., Peng, Q., DiCarlo, J., et al., 2019. smCounter2: an accurate low-frequency variant caller for targeted sequencing data with unique molecular identifiers. Bioinformatics 35 (8), 1299–1309. https://doi.org/10.1093/bioinformatics/bty790.

Xu, H., DiCarlo, J., Satya, R.V., Peng, Q., Wang, Y., 2014. Comparison of somatic mutation calling methods in amplicon and whole exome sequence data. BMC Genom. 15, 244. https://doi.org/10.1186/1471-2164-15-244.

Xu, R.H., Wei, W., Krawczyk, M., Wang, W., Luo, H., Flagg, K., et al., 2017. Circulating tumour DNA methylation markers for diagnosis and prognosis of hepatocellular carcinoma. Nat. Mater. 16 (11), 1155–1161. https://doi.org/10.1038/nmat4997.

Yang, S., Hu, T., Liu, H., Lv, Y.L., Zhang, W., Li, H., et al., 2021. Akebia saponin D ameliorates metabolic syndrome (MetS) via remodeling gut microbiota and attenuating intestinal barrier injury. Biomed. Pharmacother. 138, 111441 https://doi.org/10.1016/j.biopha.2021.111441.

Yang, Y., Zheng, D., Wu, C., Lizaso, A., Ye, J., Chuai, S., et al., 2019. Detecting ultralow frequency mutation in circulating cell-free DNA of early-stage nonsmall cell lung cancer patients with unique molecular identifiers. Small Methods 3 (10). https://doi.org/10.1002/smtd.201900206.

Ye, C., Ho, D.J., Neri, M., Yang, C., Kulkarni, T., Randhawa, R., et al., 2018. DRUG-seq for miniaturized high-throughput transcriptome profiling in drug discovery. Nat. Commun. 9 (1), 4307. https://doi.org/10.1038/s41467-018-06500-x.

You, Y., Tian, L., Su, S., Dong, X., Jabbari, J.S., Hickey, P.F., et al., 2021. Benchmarking UMI-based single-cell RNA-seq preprocessing workflows. Genome Biol. 22 (1), 339. https://doi.org/10.1186/s13059-021-02552-3.

Zhang, J., Li, J., Saucier, J.B., Feng, Y., Jiang, Y., Sinson, J., et al., 2019. Non-invasive prenatal sequencing for multiple Mendelian monogenic disorders using circulating cell-free fetal DNA. Nat. Med. 25 (3), 439–447. https://doi.org/10.1038/s41591-018-0334-x.

Zheng, G.X., Terry, J.M., Belgrader, P., Ryvkin, P., Bent, Z.W., Wilson, R., et al., 2017. Massively parallel digital transcriptional profiling of single cells. Nat. Commun. 8, 14049 https://doi.org/10.1038/ncomms14049.

Zhou, Q., Gampenrieder, S.P., Frantal, S., Rinnerthaler, G., Singer, C.F., Egle, D., et al., 2022. Persistence of ctDNA in patients with breast cancer during neoadjuvant treatment is a significant predictor of poor tumor response. Clin. Cancer Res. 28 (4), 697–707. https://doi.org/10.1158/1078-0432.CCR-21-3231.

Zhu, T., Liao, K., Zhou, R., Xia, C., Xie, W., 2020. ATAC-seq with unique molecular identifiers improves quantification and footprinting. Commun. Biol. 3 (1), 675. https://doi.org/10.1038/s42003-020-01403-4.

Zimmermann, J., Hajibabaei, M., Blackburn, D.C., Hanken, J., Cantin, E., Posfai, J., et al., 2008. DNA damage in preserved specimens and tissue samples: a molecular assessment. Front. Zool. 5, 18. https://doi.org/10.1186/1742-9994-5-18.

Zurek, P.J., Knyphausen, P., Neufeld, K., Pushpanath, A., Hollfelder, F., 2020. UMI-linked consensus sequencing enables phylogenetic analysis of directed evolution. Nat. Commun. 11 (1), 6023. https://doi.org/10.1038/s41467-020-19687-9.