# Ultrasensitive DNA sequencing using liquid biopsies enables precision medicine

Gustav Johansson

Department of Laboratory Medicine

Institute of Biomedicine

Sahlgrenska Academy, University of Gothenburg

UNIVERSITY OF GOTHENBURG

Gothenburg 2021

Cover illustration by Gustav Johansson.

Ultrasensitive DNA sequencing using liquid biopsies
enables precision medicine

gustav.johansson@gu.se

ISBN 978-91-8009-252-4 (PRINT)
ISBN 978-91-8009-253-1 (PDF)
http://hdl.handle.net/2077/67332

Printed in Gothenburg, Sweden 2021
Printed by Stema Specialtryck AB, Borås

Till min familj och vänner som
inspirerat till nyfikenhet

# Abstract

Liquid biopsies are minimally invasive and allow repetitive sampling of body fluids. Analysis of cell-free tumor DNA in liquid biopsies can be used as a biomarker for cancer. However, in most clinically relevant liquid biopsies, cell-free DNA is present at low concentrations, contains minute tumor allele frequencies, and is highly fragmented. Analysis of immune cell DNA in liquid biopsies can be profiled to examine the immune cell repertoire. However, this application requires unbiased quantification and accurate sequencing in an incredibly diverse DNA background. The overall aim of this thesis was to develop ultrasensitive sequencing approaches that enable the detection and quantification of individual molecules and single cells in these applications. We applied SiMSen-Seq, an ultrasensitive sequencing strategy based on unique molecular identifiers that enables error-free and quantitative sequencing. First, we showed that the amounts of plasma and DNA, number of targeted somatic variants, assay length, and target sequences affect the sensitivity of ctDNA analysis. We developed multiple quality control steps to evaluate a preanalytical workflow to analyze the amount of amplifiable DNA, degree of cellular contamination, and enzymatic inhibition. In patients with gastrointestinal stromal tumors, cell-free tumor DNA correlated with risk classification, treatment response, tumor size, and cell proliferation. Our data indicate that our method can be applied to monitoring treatment efficacy and identifying relapse early, especially in high-risk patients. Finally, we developed a targeted and ultrasensitive immune repertoire sequencing method to profile T-cell clonality. By studying the DNA of γδ T cells, we demonstrated that our approach is characterized by a wide dynamic range and high reproducibility and can be applied to enriched and non-enriched cells. In conclusion, we developed two flexible and simple liquid-biopsy applications that use ultrasensitive DNA sequencing to monitor cancer in patients with gastrointestinal stromal tumors and profile the immune repertoire of γδ T cells, respectively. We expect that several diagnostic applications that utilize liquid biopsies will be implemented in clinical routines in the future. Further technology development and the use of diverse types of analytes will advance this field of research. Ultimately, the development and implementation of ultrasensitive liquid biopsy-based analysis will facilitate precision medicine for more patients and improve their survival and quality of life.

Keywords: Liquid biopsy, cell-free DNA, immune repertoire, next-generation sequencing, unique molecule identifier, GIST, γδ T cell

# Populärvetenskaplig sammanfattning

Med hjälp av ett blodprov, eller andra lättåtkomliga kroppsvätskor, går det att på ett icke-invasivt vis följa cellspecifika förändringar i arvsmassan som skett under en individs livstid. Vissa av dessa förändringar är fördelaktiga, så som när immunceller utvecklar förmågan att känna igen en enorm mängd olika virus och bakterier. Andra förändringar är kopplade till sjukdomar så som cancer. När celler dör läcker cellfritt DNA (cfDNA) ut i bland annat blodomloppet. Detektion av cellfritt tumör-DNA (ctDNA) möjliggör screening, diagnostisering, tidig upptäckt av återfall samt behandlingssvar. Tre utmaningar är kopplade till dessa typer av mätningar. För det första är mängden cfDNA i blodet lågt, för det andra är det kraftigt fragmenterat och för det tredje är andelen ctDNA låg. Sammantaget kräver detta metoder som kan upptäcka enskilda molekyler. I denna avhandling har vi utvecklat kvalitetskontroller och strategier som kan förenkla och optimera arbetsflödet inom ctDNA-analys. Vi har därefter studerat mängden ctDNA i patienter med gastrointestinal stromacellstumör över tid med hjälp av en extremt känslig sekvenseringsmetod. Tekniken baseras på att varje molekyl ges en unik markör vilket möjliggör att antalet tekniska fel kan minimeras. Våra resultat visar att ctDNA framförallt förekommer i högriskpatienter samt att positivt prov korrelerar med bland annat tumörstorlek, hur snabbt tumören delar sig, och svar mot behandling. Slutligen har vi vidareutvecklat samma sekvenseringsmetod för att följa enskilda immuncellers expansion med extrem noggrannhet. Vi visar att vår metod ger mellan tio och hundra gånger bättre uppskattning av antalet celler jämfört med om man inte använder sig av unika markörer för sekvensering. Vi visar också att metoden inte leder till en snedfördelad kvantifiering av olika immunceller. Att kunna detektera förändringar i DNA med minimalt invasiva metoder skapar möjligheter som kan leda till effektivare sjukvård och förbättrad hälsa. Framtida tillämpningar kommer sannolikt ta vara på flera olika komponenter av kroppsvätskor samt mäta ett större antal förändringar vilket kommer att öka mängden information som går att få ut från ett enstaka patientprov. Med ökad känslighet och fler användningsområden kommer också nya utmaningar i att avgöra när testning är lämpligt och kan leda till förbättrad hälsa och överlevnad.

# List of papers

This thesis is based on the following studies, referred to in the text by their roman numerals.

I. Johansson, G., Andersson, D., Filges, S., Li, J., Muth, A., Godfrey, T.E. and Ståhlberg, A. *Considerations and quality controls when analyzing cell-free tumor DNA.* Biomolecular detection and quantification*, 2019*; 17: 100078.

II. Johansson, G., Kaltak, M., Rîmniceanu, C., Singh, A.K., Lycke, J., Malmeström, C., Hühn, M., Vaarala, O., Cardell, S. and Ståhlberg, A. *Ultrasensitive DNA Immune Repertoire Sequencing Using Unique Molecular Identifiers.* Clinical Chemistry, 2020; 10: 1-10.

III. Johansson, G., Berndsen, M., Lindskog, S., Österlund, T., Fagman, H., Muth, A. and Ståhlberg, A. *Patient specific monitoring of cell-free tumor DNA in the surgical treatment of patients with gastrointestinal stromal tumors.* (Manuscript), 2021.

# Content

# Abbreviations

| | |
|---|---|
| CDR3 | Complementarity-determining region 3 |
| cfDNA | Cell-free DNA |
| ctDNA | Cell-free tumor DNA |
| GIST | Gastrointestinal stromal tumor |
| MRD | Minimal residual disease |
| NGS | Next-generation sequencing |
| qPCR | Quantitative polymerase chain reaction |
| SiMSen-Seq | Simple multiplexed PCR-based barcoding of DNA for ultrasensitive mutation detection by next-generation sequencing |
| TCR | T-cell receptor |
| TKI | Tyrosine kinase inhibitors |
| UMI | Unique Molecule Identifier |
| V, D, J | Variable, Diversity, Joining |

# Introduction

The human genome project began on October 1st, 1990, aiming to sequence the human genome. It took thirteen years and cost 2.7 billion dollars [1]. Today the same analysis costs less than $1000 and takes a few days to complete. This development has provided us with the ability to detect genomic changes between individuals and helped us to, among other things, understand drug efficiency and the origin of genetic diseases [2]. Further advancements have allowed us to detect genetic changes that occur in individual cells during our lifetime. These changes to our cells' DNA are called *somatic variations* and are defined as any change that affects cells other than a gamete, germ cell, or gametocyte. Somatic variation has multiple implications in health and disease. For example, cancer occurs as a consequence of accumulated somatic mutations transforming normal cells into malignant cells that divide uncontrollably. Somatic mutations have also been implicated as having a role in aging and neurodegenerative diseases [3,4]. Another case of somatic variation is our immune system that undergoes genomic scrambling to generate a highly diverse defense system. Studying somatic variation demands an ability to detect extremely low-frequency variations different from the average cell.

The clinical benefit of being able to follow somatic mutations cannot be understated. In cancer, it allows for screening, diagnosis, and prognosis. It can be used to monitor treatment efficiency, detect minimal residual disease and relapse. It can also be used to detect treatment resistance and guide the steps of patient management [5–8]. Studying the somatic variation in the immune system could detect current or past infections, identify disease-associated autoimmune lymphocytes, or be used to monitor the immune response interaction with a tumor.

## Liquid biopsy

Tissue biopsy is a process of removing a small sample of tissue using surgery or small-needle aspiration to be analyzed in a laboratory. When performed correctly, a tissue biopsy is a precise and sensitive procedure to diagnose and confirm diseases such as cancer, infections, and organ rejection after transplantations. Unfortunately, attaining a tissue biopsy from many organs is sometimes a complicated and invasive procedure. Tissue biopsy is associated

with adverse side effects such as infections and bleeding. In rare cases, tissue biopsy has even been associated with the cancer spreading by disseminating the tumor [9,10]. It can also be difficult to localize the diseased tissue's exact position, which may result in failed biopsy or a false negative result [11]. Still, tissue biopsy has had immense importance in managing cancer and improved the healthcare of millions. However, suppose the limitations could be avoided and, in some circumstances, replaced or complemented with a less invasive procedure. In that case, the information provided by a biopsy could benefit even more patients.

A liquid biopsy is one such alternative approach; it denotes the sampling of any bodily fluid and includes blood, urine, cerebrospinal fluid, saliva, stool, and more [12]. Liquid biopsies are generally less invasive than their corresponding tissue biopsy. They contain residues from multiple tissues and allow for sampling when the diseased tissue is inaccessible, spread out, or in an unknown location. Therefore, liquid biopsy can potentially better capture the spatial and temporal heterogeneity associated both genetically and phenotypically with diseases such as cancer compared to a corresponding tissue biopsy [13,14]. There are also multiple occasions where tissue biopsy is not an available alternative, for example due to costs or if the patient's general state is too weak to motivate the invasive procedure. The major drawback with liquid biopsy approaches is when the analytes of interest are low concentration and require ultrasensitive analysis to be detected, and that some biomarkers cannot be assessed in the liquid phase, such as a tumor's morphology.

A liquid biopsy contains multiple fractions that can be used in downstream analysis. In blood, the plasma fraction contains cell-free DNA and RNAs, extracellular vesicles, metabolites, and proteins. The cellular fractions comprise white and red blood cells, platelets, and potentially disease-associated cells such as circulating tumor cells. Multiple analyses can be performed by utilizing both these fractions, increasing the diagnostic potential [12]. This thesis explores cell-free DNA (cfDNA) and immune repertoire of T-cells extracted from the plasma and the cellular fraction of a blood sample, respectively.

## Cell-free DNA

DNA can be released from cells into the circulation through apoptosis, necrosis, and active cellular secretion [15]. The majority of cell-free DNA (cfDNA) in healthy individuals has a length of around 166 bp with characteristics reminiscent of DNA extracted from apoptotic cells [16,17]. The size corresponds to the DNA wrapped around a nucleosome twice plus a 20 bp linker attached to histone 1, as shown in figure 1. Cell-free DNA can sometimes be seen in a ladder pattern corresponding to the length of DNA wrapped around two or more nucleosomes [18–20]. Depending on the origin, release mechanism, and other unknown processes, cfDNA can be shorter and longer than 166 bp. For example, in some cancer patients, cfDNA might be several kilobase pairs, which indicates that the DNA came from necrotic cell death [16]. Also, cfDNA with origin from solid tissues is usually shorter than the majority of cfDNA derived from hematopoietic cells [21,22]. Therefore, size-based selection of cfDNA can increase the sensitivity of analysis in some circumstances [23]. When cfDNA is released, it is quickly degraded through nuclease activity [24] and cleared by the liver [25], spleen [26], and kidneys [27]. The half-life of cfDNA is short and reflects the current cellular degradation in the body. In blood, the half-life is estimated to be between ten minutes and two hours [28]. The level of cfDNA in the blood is a poor diagnostics biomarker as external factors such as exercise [29], surgery [30], age [31], trauma injury [32], inflammation [33], and obesity [34] can influence the levels of cfDNA in a sample.

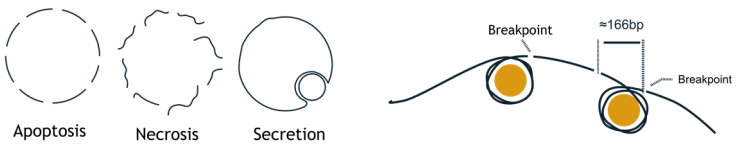


*Figure 1. Release and degradation of cell-free DNA. The left image shows three pathways for how DNA is released from cells into circulation. The right image shows how DNA is wrapped around nucleosomes (yellow), protecting the DNA, leading to DNA being cut into approximately 166 bp long segments.*

More success has been found in analyses that have investigated molecular markers in cfDNA specific to the tissue or disease of interest. In non-invasive prenatal screening, assays are specific for a fraction of cfDNA that comes from the fetus [35,36]. In transplantation, graft-versus-host reaction assays detect the fraction of DNA from the transplant (Y. M. Lo et al. 1998). In cancer, assays

target tumor-specific DNA mutations or alterations [37,38]. There are four main challenges with cfDNA analysis. First, the level of cfDNA in plasma is low and varies around 10 ng per ml in a healthy individual [39,40]. One average human diploid genome weighs approximately 6.46 pg, suggesting that each nanogram contains 310 haploid genome equivalents [41]. A milliliter of plasma, therefore, contains only a few thousand molecules upon which to base the analysis. Secondly, the analyte of interest, such as a somatic alteration, can be in concentrations lower than 1%, requiring the analysis to detect single molecules. Thirdly, cfDNA can contain background allele variants from clonal hematopoiesis and non-disease-associated mutations due to high age or benign neoplasm [42–44]. Lastly, cfDNA is highly fragmented and derived from a complex matrix such as plasma that can introduce issues in downstream analysis.

To isolate cfDNA, whole blood is separated into either plasma or serum. After blood draw, it is essential to avoid cell lysis as it risks releasing unfragmented cellular DNA, diluting the original cfDNA. Therefore, plasma is preferred over serum due to a lower risk of contamination with cellular DNA [45]. To avoid cellular degradation, plasma should be isolated within two to six hours after collection [46,47]. A second centrifugation of the plasma at high speed can remove any remaining cellular debris [48]. Plasma is subsequently stored at –80°C or directly used for cfDNA extraction. Freeze-thaw cycles of both plasma and extracted cfDNA should be avoided as it leads to DNA degradation [49]. In case plasma cannot be isolated soon after blood is drawn, preservative tubes can be used to inhibit cell lysis and nuclease activity in the sample, allowing the sample to be stored at room temperature for days [47,50]. Finally, cfDNA is extracted by methods either binding the DNA to magnetic beads or capturing it on silica-based membranes [51].

## Cellular DNA

Many of the challenges of analyzing cfDNA are not present when analyzing DNA from the cellular fraction of a liquid biopsy. In contrast to cfDNA, extraction protocols are straightforward, cellular DNA is stable, and the risk of cellular DNA contamination does not exist. Sometimes there is even a possibility to sort the cells based on cell surface markers before DNA extraction. Still, in blood cancers and in applications such as circulating tumor cells, the fraction of tumor-specific cells can be extremely low. In applications

such as immune repertoire sequencing, the proportion of DNA from a specific subpopulation like γδ-T cells can be as low as 1–5 % [52]. Detection of low-frequency clones (0.1–1 %) in these populations then requires substantial amounts of DNA, introducing a different set of technical challenges and considerations.

## Detection of somatic variants using molecular analysis

Molecular analyses of nucleic acids can be used to detect specific DNA sequences. Polymerase chain reaction (PCR) uses oligonucleotide primers specific to the sequence of interest, DNA polymerase, deoxynucleoside triphosphate, and temperature cycling to copy a selected amplicon [53–55]. Quantitative PCR (qPCR) improves on PCR and can be used to quantify the amount of DNA in a sample relative to a standard or other sample. The qPCR reaction emits fluorescence at each cycle proportional to the amount of DNA in the sample [56]. By determining at which cycle the fluorescence reaches a defined threshold, it is possible to compare the amount of target DNA at the start of the reaction [57]. By designing two sets of assays, one specific for a somatic mutation and one for the wildtype sequence, it is possible to use qPCR to determine the frequency of a somatic mutation in a sample. The assay's specificity can either be placed in the primers, using two sets of primers, or using a single set of primers and two molecular probes.

Digital PCR is a technology that can increase sensitivity and quantification even further. The principle behind digital PCR is to compartmentalize the reaction such that each target DNA molecule is amplified in a unique partition. The reactions can occur in oil droplets or small compartments in a matrix containing all reagents needed for the PCR [58,59]. Using sequence-specific probes and counting the partitions in which a successful amplification has occurred, this strategy enables digital quantification of somatic mutation on a linear scale without a standard curve. A significant limitation with PCR, qPCR, and digital PCR is that only a few targets may be differentiated at once [60]. High-throughput DNA sequencing can be used to solve this limitation and identify a wide range of mutations in a single reaction.

Sequencing can be applied to single amplicons, multiple genes, whole exomes, or even whole genomes. Multiple technologies can be used to sequence DNA. Due to the inherent properties of these technologies, not all are suitable for

detecting low-frequency somatic variants. For example, Sanger sequencing, a first-generation sequencing technology, performs a bulk analysis of all molecules in the sample using gel electrophoresis. The approach has high fidelity and makes Sanger sequencing the gold standard of sequencing. However, the method has low throughput and struggles to identify subpopulations of molecules, with a frequency below 15 to 20 %. The low sensitivity made it impractical for the detection of somatic mutations [61].

Next-generation sequencing (NGS) platforms such as Illumina, Ion Torrent, and formerly also 454 and Solid, use parallel short-read sequencing of millions of DNA molecules. This approach is better suited to separate low-frequency somatic mutations as molecules are sequenced individually instead of in bulk. In practice, NGS can reliably detect somatic mutations with allele frequencies down to approximately 1 % [62] and is limited primarily due to errors from library preparation [63], enrichment PCR, and sequencing [64]. Several ultrasensitive sequencing technologies have been developed to increase sensitivity by several orders of magnitude, as discussed in detail below.

## Error-correction in sequencing

Each region of DNA must be sequenced thousands of times to detect low-frequency mutation, a process known as deep sequencing. For example, if the variant allele frequency is 1%, only 10 out of 1000 reads will contain the mutation. However, like most scientific analyses, NGS is also limited by the signal-to-noise ratio, e.g., sequence errors (noise) will hide the true mutation (the signal). Computational and biochemical strategies can be used to reduce the number of errors in NGS. Briefly, computational methods involve filtering of data based on read-quality scores [65], analyzing the position of the error inside the read [66], and confirming the error using both read orientations [67]. Other computational methods efficiently remove adapter and primer regions, which reduce errors caused by improper alignment [68]. Furthermore, modeling the error profiles can be used [69,70] to identify particular erroneous mutation patterns, such as oxidative damage and polymerase-specific patterns [71,72], to increase confidence before calling variants.

Biochemical strategies to suppress errors involve utilizing high-fidelity polymerases in library construction. It is also important to reduce DNA damage during sample handling, as polymerases are more likely to incorporate

erroneous bases when encountering damaged DNA [73,74]. Extensive heating, ultrasonic shearing, and formalin fixation should also be avoided [75,76]. However, not all errors can be prevented or corrected. Even with the described strategies, standard NGS fails to universally call somatic mutations below 1 % [77]. In order to decrease errors even further, a new strategy was needed. Molecular barcoding introduced a simple approach to form a consensus out of a group of sequencing reads traceable back to an original molecule, discussed in greater detail below. The strategy reduces the number of errors by several orders of magnitude and could be preferably used in combination with other error reduction techniques discussed above.

## Molecular barcoding

Digital sequencing, single-molecule consensus sequencing, tag-based error correction, or molecular barcoding are all different names for similar approaches (Figure 2). By tracing and comparing 'daughter' molecules to the original 'mother' molecules, all sharing the same barcode, polymerase-induced errors that occurred during amplification and sequencing can be bioinformatically removed [64]. Detecting low-frequency mutations requires thousands of molecules to be sequenced. In molecular barcoding methods, even deeper sequencing is needed as multiple copies of each original molecule need to be sequenced to enable error correction.

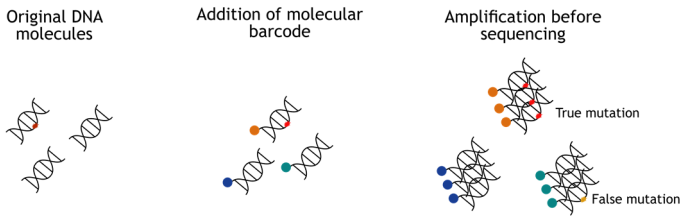


*Figure 2. Principle of molecular barcoding. Original DNA molecules, one with a mutation (red star), are tagged with a unique molecular identifier (colored dots). In library construction, reads are amplified. Errors are introduced both during amplification and when the molecule is sequenced (yellow star). Errors are distinguished from mutations as errors are only present in a subset of the molecules tagged with the same unique molecular identifier.*

There are multiple different molecular barcoding protocols; a few of them have been summarized in Table 1. These methods differ in how the barcode is attached, the barcode's structure, and how target-enrichment is performed, if needed. The barcode design can be endogenous, meaning that it is inferred

from the fragmentation or the random initiation of amplification. It can also be exogenous, meaning that it was added to the original molecules. An exogenous barcode is often a random or semi-random sequence of nucleotides called a Unique Molecule Identifier (UMI) [64]. If there are four possible bases at each position in a UMI, the number of potentially unique barcodes are four to the power of the length of the barcode. Therefore, a twelve-nucleotides-long barcode generates about 16.7 million combinations. For both endogenous and exogenous barcodes, the number of available random sequences must exceed the number of starting molecules by a few orders of magnitude. Some methods therefore utilize a combination of endogenous and exogenous barcodes to maximize the diversity. If the diversity is too low, two or more original DNA molecules may receive the same barcode and be misclassified as having a common origin. This misclassification impairs quantification and risks to falsely remove true mutations as sequencing errors [78,79].

Barcoding methods also differ in their ability to utilize one or both strands of the DNA for error correction. DNA is a double-stranded molecule and true somatic mutations, especially those with biological relevance, should be present on both strands. If a mutation is only detected on one strand, it is likely an error from sequencing or sample preparation. In methods such as duplex sequencing, each strand receives the same barcode but is error corrected independently. This approach decreases sequencing errors further, with the drawback of requiring increased sequencing depth and a complex protocol.

*Table 1 Molecular barcoding methods*

| Method | UMI attachment | Target selection | UMI structure | Reference |
|---|---|---|---|---|
| **Safe-Seq (endo), UMI-tailed Seq** | Ligation | Capture | Endogenous and exogenous | [80,81] |
| **Duplex-Seq** | Ligation | Capture | Endogenous and exogenous dual 12nt barcode | [82] |
| **INC-Seq, Circle-Seq** | None | PCR | Endogenous in vector | [83,84] |
| **Cypher-Seq** | In vector | PCR | Exogenous 7nt barcode | [85] |
| **Safe-seq (exo), SiMSen-Seq, UMI-Seq, CleanPlex** | PCR | PCR | Exogenous 12-14nt barcode | [80,86,87,87,88] |

A barcode can be added through ligation- or PCR-based strategies. A PCR-based strategy is often more efficient and may also provide target enrichment in the same step, making the preparation simpler. A ligation-based protocol is often time-consuming, might involve complicated cleaning steps, and could

lead to material losses. A ligation-based protocol requires capture-based enrichment strategies applied either before or after amplification, such as solid-phase arrays [89], RNA baits [90,91], or DNA probes [92]. However, an advantage is that it allows for broad coverage of uninterrupted DNA stretches, while PCR can struggle with regions where primers are forced to overlap [93]. PCR-based strategies may require more optimization than ligation-based approaches as primer-dimer, and non-specific amplification needs to be avoided. Also, in applications such as cfDNA where the DNA is fragmented, if the PCR amplicon cannot be kept short, ligation efficiency might be higher than for targeted PCR [94]. In other applications, such as immune repertoire sequencing where the DNA is intact from the start, ligation that requires the DNA to be fragmented should be avoided as it risks introducing unnecessary breakpoints in the region of interest and decreasing sample diversity.

One of the main challenges with ultrasensitive sequencing methods is expensive and complicated library construction protocols. In applications utilizing targeted PCR, the random sequence in the UMI contributes to primer-dimer formation and generation of unspecific PCR products. Ligation-based methods have other challenges, including time-consuming and complex library construction and low efficiencies in ligation and target DNA capture. In summary, qPCR, digital PCR, sequencing, and ultrasensitive sequencing have different advantages and disadvantages regarding sensitivity, target coverage, cost, and simplicity (Figure 3). Quantitative PCR is the most straightforward, least expensive technology. Digital PCR has increased sensitivity but is more complicated and costly. Sequencing has higher coverage and cost but lower sensitivity, and ultrasensitive sequencing has high sensitivity but can be expensive and complicated. Coverage of ultrasensitive sequencing is dependent on budget, as it requires ten to hundreds of times more sequencing capacity than traditional sequencing.

This thesis utilizes the ultrasensitive sequencing method named Simple multiplexed PCR-based barcoding of DNA for ultrasensitive mutation detection by next-generation sequencing (SiMSen-Seq) [86]. As shown in Figure 3, the SiMSen-Seq aims to make ultrasensitive sequencing simple and cost-efficient, reducing the drawbacks with current ultrasensitive solutions.

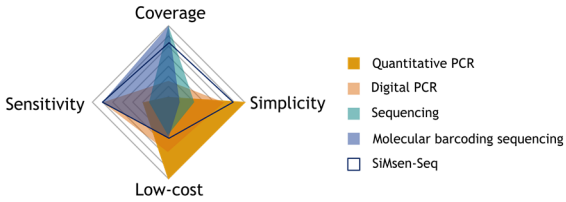


*Figure 3. Radar diagram of the performance of different molecular techniques detecting somatic variants. Sensitivity refers to analytical sensitivity. Coverage refers to the number of variants possible to cover in a single reaction. Simplicity is a combination of time, number of steps, and knowledge, required to complete the analysis. Low-cost is the combined cost of reagents, time, and necessary steps such as sequencing. A higher cost has a lower value on the low-cost axis.*

## SiMSen-Seq

The SiMSen-Seq method consists of two rounds of targeted PCR (Figure 4A). In the first round, all targeted DNA is barcoded, and in the second round, the product is amplified with sample-specific indexes to generate Illumina-compatible sequencing libraries.

The barcoding step includes three cycles of amplification where three strategies are utilized to reduce the amount of non-specific product formation. Firstly, the method uses a unique temperature-dependent hairpin loop to shield the 12-nucleotide-long UMI (Figure 4A). Secondly, a low primer concentration is used, which is compensated by extended annealing time. Thirdly, after the preamplification, the PCR is quickly attenuated and diluted by adding a TE buffer supplemented with protease. Each original DNA molecule produces, on average, six uniquely barcoded and amplifiable copies (Figure 4B). In the second step, a third of the reaction, on average two barcoded molecules per original molecule, is amplified using Illumina sequencer adapters, constructing a sequencing library.

The multiple steps undertaken to avoid non-specific product formation eliminate the need for intermediate purification between UMI tagging and library amplification [95]. As discussed above, sufficient sequencing depth must be used so that each barcode is sequenced multiple times. Sequencer data are processed through a bioinformatical pipeline. Briefly, the sequencing reads are aligned to the human genome. The reads aligning to the same location are grouped into families based on the barcode sequence. The reads within a barcode family form a consensus sequence. Due to the few cycles of

preamplification in the first round of PCR, it is also possible to accurately estimate the number of DNA molecules used to construct the sequencing library.

All barcoding methods are limited in their ability to correct errors occurring before or in the process of adding barcodes, and in that errors arising in the barcode may falsely categorize reads as novel families and therefore count them as new molecules. Bioinformatical pipelines can adjust for the second error by merging barcode families with less than one mismatch in the barcode sequence and where one of the families is considerably smaller than the other. A family size cut-off, such as three or ten reads, can be used to ensure that each molecule has been error corrected [95].

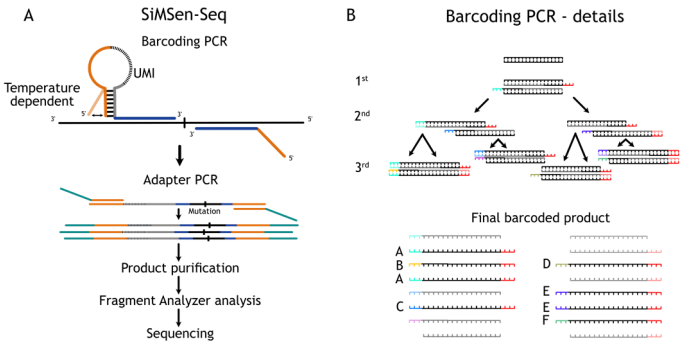


*Figure 4. Overview of the SiMSen-Seq reaction. (A) From barcoding to sequencing SiMSen-Seq consists of five steps: Barcoding PCR, Adapter PCR, product purification, Fragment Analyzer analysis, and sequencing. Target specific primers (blue), adaptor sequences (orange), SiMSen-Seq stem (grey), unique molecular identifiers (UMI, dashed line), Illumina adaptors with P5, P7, and index (turquoise). (B) A detailed schematic of the three cycles of amplification in the barcoding step. Amplification starts from a targeted primer tagged with a unique molecule identifier (coloreds ends) and a targeted primer containing only the adapter sequence (red end). DNA synthesized in the 1st, 2nd, or 3rd cycle is indicated as translucent. The final barcoded product will consist of six uniquely barcoded molecules. Molecules marked with A and E in two copies, and molecules marked with B, C, D, F in one copy*

The SiMSen-Seq method has been used to detect ctDNA in esophageal cancer [96], melanoma [97,98], colorectal cancer [99,100], and head and neck cancer [101]. It has been used to detect mutations in the cellular fraction of bone marrow and PBMC when monitoring hematological malignancies in humans [102,103] and in mice [104]. It has also been used for basic research to study polymerase fidelity [105], UV-induced damages [106], and genetically

modified plants [107]. A multilaboratory assessment showed that SiMSen-Seq, in contrast to other comparable methods, reliably detected samples with 0.125 % variant allele frequency. [108]. In conclusion, multiple research groups have demonstrated SiMSen-Seq as a flexible and easy-to-use ultrasensitive sequencing method.

## Circulating tumor DNA

As discussed earlier, liquid biopsy can be used in cancer management to detect circulating tumor DNA (ctDNA) and has become a powerful biomarker predicting poor patient outcomes and has supported personalized medicine [7,20,109]. Levels of ctDNA correlate with tumor volume, stage, and disease burden [110–112].

Detection of ctDNA has different strengths and weaknesses in the various stages of cancer management. Screening based on ctDNA has the advantage that it does not expose the patient to radiation such as computed tomography, is a minimally invasive procedure, and allows sample collection at primary care. As a screening test, the sensitivity on a population level is dependent on how often the test is performed. It is therefore essential that cost is kept low. Larger panels are costly, whereas narrow panels will only capture a subset of all cancers. Notably, if the sequencing panel does not cover cancer-specific mutations or alterations, even a patient with a high disease burden will receive a false negative diagnosis. A negative ctDNA analysis should, therefore, never be used to rule out cancer; however, for several cancers that currently have no screening option, using ctDNA to capture some patients before clinical onset could have clinical value. Especially in specific risk populations, such as older heavy smokers, lung cancer screening with narrow panels might be a feasible approach [113].

Increased screening and sensitive analyses may lead to overdiagnosis and overtreatment [114]. Of all mutations in cfDNA found in healthy controls and cancer patients, about 80% and 50%, respectively, likely arrived from clonal hematopoiesis [42] and not from any malignancy. There is also a risk of identifying cancers that do not motivate treatment and, if detected, would only increase anxiety. The challenges with benign somatic alterations and tumors that do not require management in cfDNA analysis will also increase with age [44].

When cancer is detected, ctDNA can be used as a prognostic and predictive biomarker, comparable to tissue biopsy, for genomic characterization of the tumor [7]. As previously discussed, a liquid biopsy is less invasive, quicker, and more cost-effective than a tissue biopsy [115]. Liquid biopsy is also potentially better at capturing the spatial and temporal heterogeneity of the tumor. However, tissue biopsy still has higher clinical sensitivity, especially for small tumors [116], and might also add other biomarkers beyond genomic characterization, such as histology. In applications such as managing *EGFR*-positive lung cancer, ctDNA analysis is therefore only offered as an alternative when a tissue biopsy is not achievable or as a complementary analysis [117].

Before, during, and after treatment of confirmed malignancy, routine ctDNA analysis allows for monitoring treatment efficiency and detecting minimal residual disease and relapse [117–120]. It could also enable early detection of mutations associated with treatment resistance allowing the physician to change therapeutic strategy.

## Precision medicine

Precision medicine, sometimes referred to as personalized medicine, is commonly used to tailor medical treatment to a subset of patients, often carrying specific genetic markers. Precision medicine enables interventions to be focused on the patients who will benefit, avoiding side effects and costs for those who will not [121]. During the recent decade, genetic alterations in tumor DNA have been increasingly used to guide treatment in cancer patients. Mutation analysis in tissue biopsy is currently the gold standard to detect these genetic markers, but liquid biopsy could increase the number of available patients for personalized medicine approaches as test can be performed at more circumstances. Liquid biopsy-based precision medicine has found most application in cancers such as lung, melanoma, colon, breast, and prostate cancer where there is a strong correlation between genetic markers and treatment efficiency [122].

In managing metastatic non-small-cell lung cancer, the National Comprehensive Cancer Network recommends measurements of a minimum of nine biomarkers in the genes *EGFR, ALK, ROS1, BRAF, RET, MET, HER2,* and *NTRK*. For example, mutations in *EGFR* make this type of lung cancer sensitive to EGFR tyrosine kinase inhibitors and occur in around 10% of all

cases [123]. Unfortunately, 60% of patients acquire resistance towards first line of treatment within 9 to 10 months due to a T790M mutation in *EGFR*. Subsequently, these patients are currently treated with osimertinib, which irreversibly inhibits the EGFR despite the T790M mutation [124]. In melanoma, mutations in *BRAF* are required for treatment with BRAF and MEK inhibitors [125], and *NRAS* mutations are associated with resistance to multiple drugs [126]. In metastatic castration-resistant prostate cancer, mutations and amplifications of the *AR* gene are associated with treatment resistance and could help with patient stratification [127]. In metastatic hormone-positive breast cancer, mutations in *ESR1* and *PIK3CA* could predict responsiveness to aromatase inhibitor and PI3Kα-selective inhibitor, respectively [128,129]. Lastly, in colon cancer, mutations in *KRAS*, *NRAS,* and *BRAF* indicate resistance to anti-EGFR treatment [130].

Still, there are only four FDA-approved companion diagnostics for ctDNA on the market to date: FoundationOne Liquid CDx [131], Guardant360 CDx assays [132], Cobas EGFR Mutation Test v2 [133], and PIK3CA RGQ PCR Kit [134]. The first two are based on an NGS panel for comprehensive genomic profiling detecting mutation in multiple genes, and the last two are based on qPCR and measure a selection of mutations in each indicated gene.

In this thesis, we study patients diagnosed with gastrointestinal stromal tumor (GIST). This cancer type is the most common abdominal sarcoma with a yearly incidence of 15 cases per million inhabitants [135–137], and was one of the first cancers where treatment benefited from a personalized medicine approach [138]. More than 90% of GIST tumors harbor a mutation in *KIT* or *PDGFRA* that sensitizes them to tyrosine kinase inhibitors (TKI) [139,140]. Therefore, mutation analysis became the standard of care for these patients [138]. Surgery is often curative for low- and intermediate-risk group patients, while high-risk tumors are treated with TKI both before and after surgery if a sensitizing mutation is detected [141]. The personalized medicine approach of high-risk GIST patients has resulted in a significant increase in disease-free and overall survival [142]. However, despite the absence of detectable tumor after surgery, most high-risk patients experience recurrence and primary or secondary TKI resistance after five years [143]. The connection between tumor genomics and available precision medicine argues for the potential of utilizing ctDNA as a

biomarker to monitor treatment efficiency, recurrence, and the development of resistance mutation in GIST [144–147].

## Immune repertoire sequencing

Another biomarker with potential in cancer management and other diseases is monitoring T and B cells' immune repertoire. These cells undergo a remarkable alteration of their DNA during cell maturation, generating the diversity found in our immune systems to detect and react to all available antigens [148,149]. The T-cell receptor (TCR) can either be αβ encoded by the TRA and TRB locus or γδ encoded by TRD and TRG locus. The B cell receptor is produced similarly but is formed from a heavy chain locus (IGH) and two light chains loci (IGK and IGL). Each locus contains numerous variable (V) genes and joining (J) genes, and some loci also contain diversity (D) genes [149]. The TRD locus is studied in this thesis due to its implicated role in multiple sclerosis [150,151], is located on chromosome 14, and contains eight V-, four J-, and three D-genes (Figure 5). Immune recombination is a complex process. Briefly, during maturation of the immune cell, the RAG1 and RAG2 proteins bind and cleave the DNA to select and join one V-, D -, and J-segment semi-randomly. [152]. In addition to this selection of gene segments, terminal deoxynucleotidyl transferase is used to delete and add random nucleotides between the joined pieces [153]. Combining these two processes creates an enormous diversity in the complementarity-determining region 3 (CDR3), which is the critical part of the δ chain that makes the receptor specific for antigens.

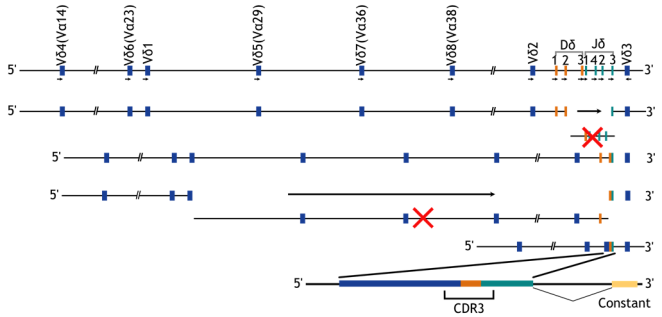


*Figure 5. Overview of VDJ recombination. The TRD locus contains V genes (blue), D genes (red), and J genes (green). Arrows show transcription direction. In multiple intermediate steps (not shown), the DNA is recombined to join a random V, D, and J segments. The transcribed product is spliced to join a constant region (light yellow) used in some RNA-based immune repertoire sequencing applications.*

The CDR3 sequence is inherited when a T or B cell divides, so by sequencing the CDR3 region, the clonal expansion of specific T and B cells can be monitored. Immune repertoire sequencing has been used to study the immune system in multiple applications, such as vaccine development [154], autoimmune disorders [155], and cancer [156]. It can be used in cancer management to monitor minimal residual disease in lymphoma and leukemia [157] and predict prognosis by characterizing tumor-infiltrating lymphocytes [158]. Immune repertoire sequencing could also monitor the direct effect of immune checkpoint therapy by detecting changes in the repertoire diversity [159] and tracking the development of adverse advent associated with these therapies [160]. ClonoSEQ, a targeted NGS assays for immune repertoire sequencing, recently became the first FDA-approved NGS assay for minimal residual disease in chronic and acute lymphocytic leukemia and multiple myeloma [161].

Sequencing error, biased amplification, and quantifying the number of analyzed cells are three significant challenges in immune repertoire sequencing [156]. Sequencing errors make it difficult to separate low-frequency sequence-similar clones from erroneous base calls and lead to an inflated diversity [162,163]. Unbiased amplification leads to biases in quantifying clones' subtypes and the number of cells included in the analysis. The aforementioned strategy to correct these types of errors is to use UMI. So far, UMI in immune repertoire sequencing has mainly been used on mRNA [162–164]. However, RNA transcription levels vary per cell [165,166], and reverse polymerases are more prone to errors [167,168] and have variable efficiency depending on sequence [169]. The advantage of mRNA-based approaches is that primers can amplify from the so-called "constant" region joined after splicing downstream of the CDR3 in the mRNA transcript (Figure 5). This strategy reduces amplification bias as one constant primer can be used instead of a set of different J-primers [170]. Still, DNA-based methods are preferable if accurate quantification of cells is the experiment's main objective. As discussed earlier, UMI can be added either through ligation- or PCR-based approaches. After UMI attachment, all fragments are amplified using the same universal adapter primers. In this thesis, we developed the first PCR-based ultrasensitive sequencing approach for immune repertoire sequencing. Previously only ligation-dependent approaches were available [171]. Recently, additional PCR-based methods utilizing UMI for immune repertoire sequencing have been published [88].

# Aims

of liquid biopsies using ultrasensitive sequencing. Blood consists of a cellular fraction and a non-cellular plasma fraction that can be used for biomarker analysis. Here, we studied somatic variations in cfDNA and T cells. In both applications, detection and quantification of individual DNA molecules with single nucleotide resolutions are needed. To enable reliable DNA analyses, the entire workflow from sampling, via extraction and sequencing, to data analysis needs to be carefully optimized. This thesis focuses on the potentials and challenges of ultrasensitive DNA analysis using liquid biopsy.

Specific aims:

Paper I: To develop quality controls for the analysis of cfDNA, including ctDNA, in blood plasma. We also aimed to develop a framework to increase sensitivity, including sample volume, multiplexing, and assay length.

Paper 2: To develop and apply a personalized and ultrasensitive ctDNA sequencing approach to monitoring patient-specific mutations and TKI resistance in liquid biopsies from patients diagnosed with gastrointestinal stromal tumor undergoing surgery.

Paper 3: To develop an ultrasensitive immune repertoire sequencing strategy for analyzing γδ T-cell receptor clonality in healthy individuals.

# Results and discussion

## Considerations and quality controls when detecting ctDNA.

There are three major challenges when detecting ctDNA. Firstly, there is a low amount of cfDNA in plasma. Secondly, the cfDNA is highly fragmented, and thirdly, in early-stage cancer patients, the tumor allele fraction is low. Altogether, this led to few ctDNA-molecules in a sample containing a particular mutation. Therefore, ultrasensitive ctDNA analysis requires sensitive analytical techniques, such as SiMSen-Seq, but also an optimized workflow from sampling to data analysis in order to enable accurate and reliable liquid biopsy assessment. In Paper I, we explore experimental considerations and quality controls useful when performing ctDNA analysis (Figure 6).

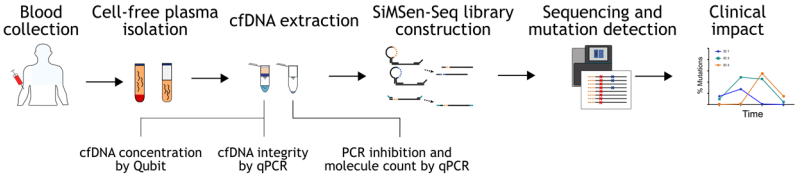


*Figure 6. The general workflow of ctDNA analysis utilizing SiMSen-Seq, including recommended quality controls.*

## Increasing sensitivity to detect ctDNA in liquid biopsy.

The sensitivity of mutation analysis using liquid biopsy is limited by the number of mutated tumor-specific molecules in a sample. A sample with a low-frequency mutation at 0.1 % requires 3.6 ng DNA to, on average, contain one ctDNA molecule with the mutation (Figure 7A). However, due to the Poisson distribution, such a test's sensitivity—assuming that it is possible to detect a single mutation—will be only 63 %. To be 95 % confident that the sample always contains at least one such ctDNA molecule, it needs to contain, on average, 4.7 molecules. Furthermore, in most applications, even using molecular barcoding methods, more than one mutated molecule needs to be detected to be confident enough to call the variation. There are two strategies to increase the number of tumor DNA molecules that can be analyzed in a sample. First, it is possible to increase the number of assays. Second, it is possible to increase the amount of DNA by increasing the volume of plasma.

If two independent assays monitor two different tumor-specific mutations, twice as many mutations can be detected, and this increases the assay's sensitivity (Figure 7B). The approach has successfully been used for detecting minimal residual disease by detecting multiple mutations confirmed from the tumor biopsy [172,173]. Notably, the approach is not applicable when a single mutation is of interest, for example, development of resistance to a particular drug. The second approach is to increase the volume of plasma extracted from the patient. Doubled plasma volume theoretically doubles the number of ctDNA molecules in the sample. Both these strategies can be used in combination. However, it will increase required sequencing and therefore cost.

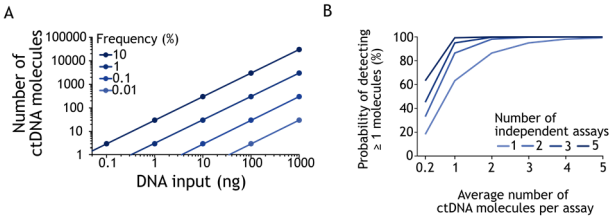


*Figure 7. Analysis of theoretical numbers of ctDNA molecules (A) Number of ctDNA molecules with the specific mutation* is *dependent on the amount of DNA and the frequency of the mutation. (B) The probability of detecting at least 1 molecule depends on the number of ctDNA molecules per assay and the number of independent assays. Adapted from [174].*

## Quality controls of cfDNA

As previously discussed, all experimental steps, such as selecting blood-collection tubes, plasma preparation, plasma logistics, and extraction method, affect cfDNA analysis. These preanalytical factors impact the yield of cfDNA, the risk of contaminating the cfDNA with DNA from post-withdrawal apoptotic or necrotic cells, and may also introduce or enrich analytical inhibitors. Yield and inhibitors affect the number of amplifiable molecules in the sample and directly impact sensitivity. Contamination of cellular DNA dilutes the ctDNA and will cause errors in estimating the mutant allele frequency.

The first quality control used in paper I was the measurement of yield after extracting cfDNA. A simple analysis can be performed using a device such as nanodrop. However, the method is sensitive to contaminants and may overestimate the concentration. In paper I, we used fluorometers as an alternative and more accurate approach [45]. Both these methods detect the

total amount of DNA in the sample. However, in targeted sequencing, not all DNA will be available for amplification as the primer's binding regions may fall outside the DNA fragments (Figure 8). The theoretical percentage of cfDNA molecules that can be amplified in targeted PCR, assuming an average fragment length of 166 bp, can be calculated as $1 - (n/166)$, where n is the amplicon's length [94]. This formula suggests that a 100-base-pair-long assay can amplify 40 % of the total DNA and assumes that the cfDNA is randomly fragmented. However, epigenetic factors such as nucleosome positioning have a considerable influence on cfDNA degradation. Therefore, some loci will be more degraded than others, leading to fewer amplifiable molecules [175]. In paper I, we show experimentally that amplicon length correlates with amplifiable DNA for randomly fragmented DNA but less for cfDNA (Figure 9). Therefore, qPCR utilizing the same target primers as sequencing primer will provide more accurate quantification of the number of the sequenceable molecules in the sample. Quantifying the amount of cfDNA first using fluorometers and then with qPCR using our assay of interest showed that only 49% of the cfDNA was amplifiable. It also concludes that it is essential to design short assays when analyzing cfDNA and, if possible, avoid regions prone to degradation.

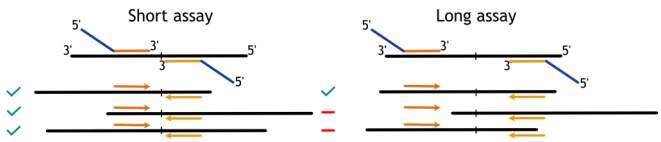


*Figure 8. Amplicon length influences the number of amplifiable molecules. When the primer binding region is outside the DNA template, the assay will not amplify.*

The second quality control used qPCR to quantitatively assess the amount of contaminating cellular DNA in a liquid biopsy. We did this by utilizing a long and a short qPCR assay. The shorter assay amplified all DNA, while the longer assay only amplified DNA that is longer than typical cfDNA and, therefore, likely contamination. The difference between these two assays provides the degree of contamination. This test can be beneficial to perform when evaluating a new workflow or testing a sample that has been stored sub-optimally and could have been contaminated with cellular DNA. In our workflow, where the plasma was collected in Norgen cfDNA preservative

tubes and then extracted using a Magmax cell-free DNA isolation kit, we detected cellular DNA in 12.5 % of all samples. Only one sample had high enough contamination to significantly impact the mutation allele frequency if a mutation were to be detected.

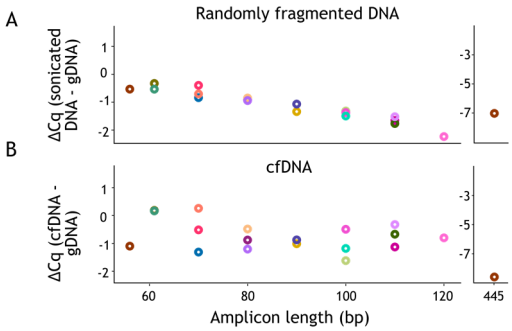


*Figure 9. Assay amplifiability depend on amplicon length. The position of each colored circle indicates the mean difference in cycle of quantification (Cq) value comparing sonicated DNA (A) and cfDNA (B) with genomic DNA (gDNA) for nineteen independent assays (n = 3). Adapted from [174].*

The third quality control was done after the sample had been concentrated. The concentration is necessary to maximize the amount of DNA loaded into the sequencing reaction but may also result in losses and concentrate inhibitors. Using qPCR, we showed that it is possible to assess sample inhibition by inspecting the amplification curve. It was possible to rescue single samples by either diluting or re-extracting the sample. If many samples are inhibited, it suggests that something is wrong with the current workflow. In paper I we showed that changing the extraction method could remove inhibition. Such issues could, for example, be due to incompatible collection tubes and extraction methods. As long as the sample is uninhibited, this final qPCR also provides accurate quantification before the sample is loaded into the sequencing reaction and could be used to calculate the required sequencing depth. We show a strong correlation between the amount of DNA loaded into the sequencing reaction based on our qPCR data and the number of barcoded molecules detected after sequencing.

In summary, the number of amplifiable molecules will be dependent on the total amount of cfDNA in the sample, the degree of fragmentation, the degree of PCR inhibition, and the degree of losses in preparing the material before construction of the sequencing library. These losses could be monitored and

hopefully minimized using quality controls through the preanalytical steps, increasing the sensitivity of the final analysis.

## Precision medicine in gastrointestinal stromal tumor

Management of patients with gastrointestinal stromal tumors (GIST) is one of the earliest examples of personalized medicine and has significantly improved overall survival [142]. In Paper II, we applied the experimental workflow and quality controls developed in Paper I and developed patient-specific SiMSen-Seq panels to monitor ctDNA. Blood plasma samples were collected during routine controls both before and after surgery. Three samples were also collected in connection to surgery at the start, during mobilization of the tumor, and at wound closure. Patients from all risk groups were included in the study. The personalized panels targeted the tumor-specific mutation, identified from routine sequencing of tissue biopsy and the most common loci for TKI resistance mutations.

This exploratory study aimed to determine how ctDNA correlated with clinical parameters, including disease risk status, tumor size, and treatment response. The study included 32 patients and analyzed 161 plasma samples. We detected ctDNA in 9 out of 32 patients; all but one were high-risk. Patients positive for ctDNA had significantly larger tumors and higher cell proliferation as analyzed with protein marker Ki-67. Interestingly, all ctDNA-positive patients had either *KIT* or *PDGFRA* insertion or deletion, and none had single nucleotide variants. The detection of ctDNA was associated with treatment response. All patients positive during surgery became negative in the sample following surgery. The study included seven patients with metastatic disease. Three were ctDNA-positive at any point in time and the detection was associated with disease progression. Of the four negative metastatic patients, three were included after TKI treatment initiation.

Only 50 % of treatment-naïve high-risk patients had detectable ctDNA, which is comparable to other studies. These results suggests that GIST is a low-shredding tumor type and that tumor-specific mutation in ctDNA analysis is not a sensitive biomarker at the diagnosis of GIST. Still, ctDNA was associated with active disease in high-risk patients. In two patients, we detected resistance mutations, in both cases this could have an impact on treatment decision. In one patient, the treatment resistance mutation was detected before surgery,

showing that monitoring could be beneficial both before and after surgery. A conclusion is that monitoring ctDNA in high-risk patients may facilitate management and has the potential to improve patient outcomes, but that a larger cohort is needed to identify the true clinical utility.

Our patient-specific ultrasensitive assays could detect variants at an allele frequency between 0.04 and 93 %. The unique molecular identifier in the SiMSen-Seq assays enables demultiplexing and error correction, as previously discussed. Figure 10A shows an example of a palliative GIST patient not included in the study as the patient did not undergo surgery. By analyzing the raw data without considering the UMI, the background was too high to call the patient-specific variation from sequencing errors. Using SiMSen-Seq error correction, only the variant consistent with the known patient-specific mutation is left, and it is possible to call the variation confidently. The example is even more extreme in Figure 10B from the same patient. Using SiMSen-Seq, it is possible to call a variation known to be associated with treatment resistance. Here the position is unknown, and it would be impossible to detect without error correction due to the high background.

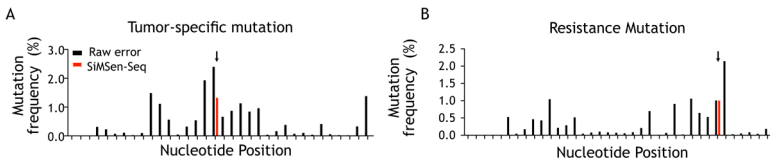


*Figure 10. Error correction using SiMSen-Seq in clinical samples. (A) Tumor-specific mutation is detected (arrow) slightly above background (black bars) when utilizing error correction (red bars). (B) In the same patient a treatment resistance mutation is detected (arrow) that would be impossible to call without error correction.*

## Calling variants in cancer applications

In paper II, a patient-specific single nucleotide variant was called if the sample contained more than six error-corrected consensus reads with the mutation. If a mutation were called with this criterion, we only required a single consensus read containing the variant for the other samples from the same patients, as the mutation could be suspected. If the mutation was an insertion or deletion of nucleotide, we also only required one single molecule to call the variant as insertions and deletions are uncommon sequencing artifacts for Illumina sequencing. Six molecules were used as a cut-off for single nucleotide variations because errors occurring in sample preparation and the first stage of

barcoding could at most give rise to six barcoded molecules. The SiMSen-Seq bioinformatical workflow corrects errors arising after barcoding. The exact cut-off had little influence on the overall results; however, this manual approach to variant calling is a weakness of paper II.

Variant calling software with more sophisticated approaches adapted for barcoded sequencing are available [176–178], but none of these methods has been validated for SiMSen-Seq datasets and was therefore not used. More generic approaches utilize tools like fgbio [179] to construct error-corrected consensus reads combined with traditional variant callers like Mutect [66] and VarDict [180]. However, these variant callers fail to detect low-frequency variants as they are adapted to deal with data containing background noise corresponding to standard NGS [178]. Interestingly, no approach to our knowledge takes user-guided information about patient-specific mutation or the common treatment resistance mutation as input to adjust the variant calling software's sensitivity and specificity.

## Future of liquid biopsy in GIST

One of the main clinical benefits of monitoring high-risk GIST patients using ctDNA is the early detection of resistance mechanisms. Patients who experience tumor-progress on imatinib can benefit from second and third-line TKIs, such as sunitinib and regorafenib, respectively [139,181,182]. At least seven different TKIs are available, and more are in development. [182]. A fascinating development are drugs that target variants of *KIT* and *PDGFRA* with already acquired therapy resistance [183]. In the future, this will provide physicians with an arsenal of therapeutics for different tumor mutations. The development is similar to the management of ALK-driven neuroblastoma, where multiple tyrosine kinase inhibitors exist for different *ALK* point mutations [184]. A hopeful clinical case in neuroblastoma suggests that a patient could rotate between all available ALK-specific treatments, eventually return to and respond to the initial first-line treatment and later become disease-free [185]. It is also essential to understand other escape mechanisms of the tumor besides those reducing specific TKI inhibitors' efficiency. Exploratory GIST studies propose new druggable targets for patients progressing after first and second-line TKI inhibitors [186] involving genes other than *KIT* and *PDGFRA*. Therefore, the number of assays required to monitor TKI resistance mechanisms will likely grow in the future.

## Ultrasensitive immune repertoire sequencing

The benefit of monitoring will only increase the more we learn about tumor development. The tumor's ability to evade our immune system is another hallmark of cancer [187]. So far, there is a lack of useful liquid biomarkers to monitor immune tolerance or predict immunotherapies' efficiencies. In the third paper, we developed an ultrasensitive method for immune repertoire sequencing that could find potential use in the monitoring of cancer and other diseases in the future.

Immune repertoire sequencing identifies and quantifies the number of T- or B-cell clones in a sample using sequencing. In Paper III, we developed a method for immune repertoire sequencing based on targeted amplification of DNA utilizing UMI to reduce sequencing errors and enable digital quantification (Figure 11). As previously discussed, the advantage of applying UMI for immune repertoire sequencing is to improve error correction and digital quantification of the number of cells analyzed, increasing confidence when detecting and quantifying low-frequency clones. We developed a proof of principle to study the rearranged TRD locus in γδ T cells. Target-specific forward and reverse primers were designed for each *TRDV* and *TRDJ* gene, respectively. In total, eight *TRDV* primers and four *TRDJ* primers were designed and used to capture the full diversity of the TRD locus. To determine each *TRDV* and *TRDJ* primer combination's efficiency, we designed 32 synthetic molecules containing the target sequence of respective TRDV and *TRDJ* genes and a template-specific sequence. We then used a standard curve of the synthetic molecules and performed qPCR on each of the 32 assays to measure the efficiency. All assays performed close to 100% efficiency.

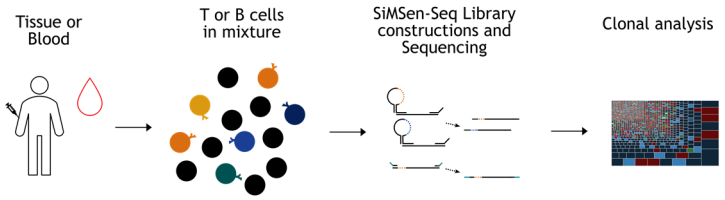


*Figure 11. Workflow for ultrasensitive immune repertoire sequencing. Blood is collected, white blood cells are isolated. An optional enrichment step can be used to purify cells of interest such as γδ T cells. A SiMSen-Seq reaction with immune repertoire primers are used to create sequencing library. Sample is sequenced, data are run through bioinformatical pipeline and clonal analysis can be performed.*

To enable ultrasensitive sequencing using SiMSen-Seq, we incorporated a 12 bp long UMI between the adaptor sequence and the 5' end of each forward primer. Our sequencing approach contained two rounds of PCR. In the first, all DNA was barcoded, and in the second, the library was amplified with Illumina index primers, like normal SiMSen-Seq. The number of barcoded molecules generated from the first round of PCR was validated using qPCR and a standard curve of synthetic molecules. Each assay provided a specific product in the dynamic range between 10 and 10 000 molecules. The specificity of the amplified product was evaluated using electronic parallel gel electrophoresis. We then assessed the final 32-plex using the same approach with a pool of the 32 different synthetic molecules. The multiplex showed a dynamic range from 20 to 20 million molecules with a PCR efficiency of 101 %. We then sequenced the libraries generated from 20, 200, and 2000 molecules to evaluate each primer-combination performance individually. Each assay had close to 100 % efficiency when analyzed by sequencing (Figure 12A). We evaluated each assay's sensitivity by reducing the synthetic molecules' concentration to approximately 10 molecules per reaction. The result indicated that each assay could detect this low concentration of target molecules in a diverse background of synthetic molecules.

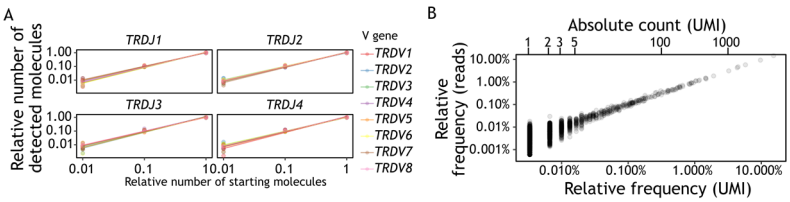


*Figure 12. Unbiased amplification and improved quantification of immune repertoire assays. (A) Individual efficiency of 32 assay combinations using a standard curve based on synthetic molecules. Values are normalized n = 3. (B) Improved quantification using UMIs. The relative frequencies of clonotypes using UMI (x-axis) versus raw sequencing reads (y-axis). Values are converted to absolute molecules count on top x-axis. Adapted from [188].*

The 32-plex were then validated on DNA extracted from enriched γδ T cells from human buffy coats. A similar level of efficiency was achieved when using a standard curve of extracted human DNA. However, some assays targeting rare combinations could not detect any molecules at low DNA concentration making efficiency calculation impossible. We utilized a bioinformatical tool called Molecular Identifier Guided Error Correction pipeline (MIGEC) for analyzing the data [162]. Briefly, the tool processes the raw sequencing reads

and groups together reads with identical UMI. An error-corrected consensus sequence is formed from each UMI family, reducing the challenge in immune repertoire sequencing of separating real clones from sequence errors. The process also removes the amplification bias introduced during library amplification and enables digital quantification of the number of molecules analyzed. We showed that the use of UMI reduced amplification biases between 10 to 100 times for low frequency clones (Figure 12B).

We further validated our sequencing method on isolated γδ T cells from 10 human blood samples. We show a strong correlation of the frequency of the two subfamilies Vd1 and Vd2, determined by fluorescent activated cell sorting and our sequencing protocol. We detect an oligoclonal expansion of clones from a healthy individual and a predominance of *TRDV2* and *TRDJ1* gene usage (Figure 13). This oligoclonality is found in most adults and is associated with cytomegalovirus infection and age; the expansion indicates an adaptive role of γδ T cells [189].

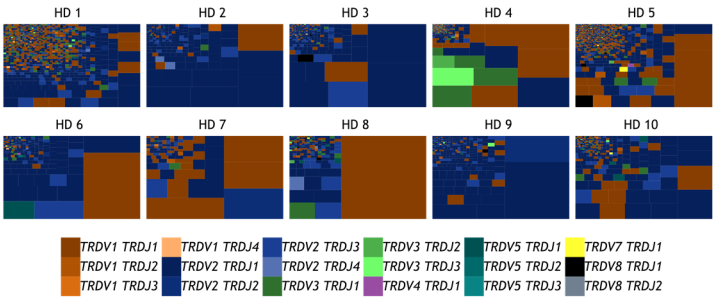


*Figure 13. Clonal distribution in ten healthy individuals. Each square represents a unique clonotype. The square area shows the clonotype frequency and the color which V and J genes were used. Adapted from [188].*

In blood samples, γδ T cells are in a low frequency, and it is not always possible to enrich before analysis. Unsorted samples result in a low proportion of amplifiable DNA and could increase non-specific binding and interfere with the PCR. To evaluate this risk, we divided the white blood cells from a blood sample in two; one part was enriched for γδ T cells, and the other was left untreated. DNA from both samples was analyzed, and the frequency of clonotypes was compared. As predicted, the non-enriched sample contained more non-specific products assessed both on parallel capillary electrophoresis and sequencing. However, both samples showed a considerable overlap of

clonotypes and a strong correlation of frequencies suggesting that the method works on enriched and non-enriched samples (Figure 14).

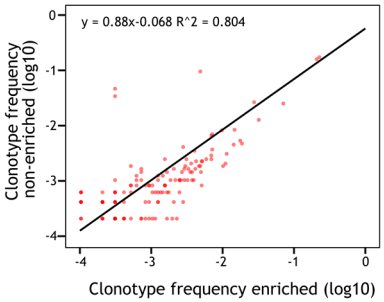


*Figure 14. Correlation of clonotypes from enriched and non-enriched γδ T-cells. White blood cells were split into two samples, and one was enriched for γδ T-cells. The frequency of all clonotypes found in both samples is shown. Adapted from [188].*

Targeted immune repertoire amplification introduces a novel set of challenges. To include coverage of rare recombination, an increasingly large primer pool needs to be designed. The inclusion of more primers increases the risk of primer-dimers without a proportional increase in target DNA. Careful optimization has previously been performed in targeted PCR approaches to perform these multiplex reactions without introducing unacceptable levels of bias or non-specific primer formation [190]. Introducing a random nucleotide sequence as a UMI increases complexity of the primer pool even further. This challenge will be more prominent when ultrasensitive immune repertoire sequencing is developed for other more complex loci. The IGH locus used in B cells, for example, contains about 50 functional V segments and six J segments [149]. SiMSen-Seq has the advantage of shielding the UMI, reducing primer-dimers, which could make these larger multiplexes feasible.

## Applications of ultrasensitive immune repertoire sequencing

One of the potential applications of immune repertoire sequencing is antibody discovery. High-affinity antibodies can be detected by sequencing the complete immune receptor with matched protein profiling. The traditional RNA and DNA-based immune repertoire sequencing solutions are limited here as they sequence each chain independently and often only part of the chain. It is possible to solve this problem, at least partly, using subsampling of cells [191]. Other approaches utilize single-cell emulsions with linkage PCR or

single-cell sequencing to match chains and construct a complete TCR or BCR sequence [192]. Single-cell sequencing can in addition also capture the cell's transcriptome to describe the cell phenotypically. These approaches require single-cell suspension and can be complicated and costly. However, possibly single-cell techniques can be complemented with ultrasensitive sequencing of bulk material when this is sufficient.

Using only the CDR3 sequence of one chain, it can be possible to predict past inflammatory responses. Recently, databases for TCR and BCR sequences with matched HLA and antigen have been created to facilitate annotation of CDR3 sequences [193–195]. From these databases, it is possible to see if patients have a CDR3 sequence associated with previous infection. Unfortunately for the γδ T cells field, only the Pan Immune Repertoire Database contains information related to γδ T cells and have considerably fewer entries for γδ T cells than for αβ T cells and B cells [195]. However, the immune repertoire is highly private, suggesting that even if two individuals react to the same antigen, the CDR3 sequences may be significantly different from each other. The "exactness-requirement" can be problematic for databases. To solve this, novel computation approaches and machine learning allow the clustering of sequences that share similar features to predict the function of specific CDR3 sequences and immune repertoires without requiring an exact match [196,197].

Immune repertoire sequencing has demonstrated superior sensitivity in detecting minimal residual disease (MRD) from bone marrow transplants in lymphoma and leukemia compared to flow cytometry and allele-specific PCR [198]. A malignant clone's CDR3 sequence is a tumor-specific sequence that can be monitored in MRD. However, in MRD application, clones' sequences are diverse enough that if one or two errors occur, a bioinformatical pipeline can correctly remove these errors. This error redundancy makes immune repertoire sequencing different from ctDNA mutation, and in concordance with the previous discussion from paper I, the sensitivity will primarily be dependent on the number of amplifiable molecules. Therefore, in MRD applications, in particular, it is pivotal that primers have high efficiency for all gene variants, capturing all available DNA. In summary, primer efficiency and complete coverage are more critical than ultrasensitive sequencing for ultrasensitive detection in MRD applications. However, as long as the

increased cost of redundant sequencing does not lead to a lower amount of DNA being used in the analysis, UMI provides improvements, such as accurate quantification.

Another immune repertoire sequencing application is vaccine development. Processes, like affinity maturation—exclusive to B cells—introduce small nucleotide variants to increase the antibody's affinity and can be challenging to separate from sequence errors. Accurate quantification of both primary and secondary responses from effector and memory T cells is also necessary to monitor in vaccine development [199]. Both these challenges benefit from the increased sequencing accuracy and improved quantification offered by UMI-based immune repertoire sequencing.

# Conclusion

Liquid biopsies are minimally invasive and enable repetitive monitoring of patients. Sequencing of DNA in both the cellular and non-cellular fraction can be used to study somatic variations, including ctDNA and the immune repertoire. The sensitivity for detecting somatic variations requires both an optimized workflow that minimizes material losses and an ultrasensitive analytical approach. Ultrasensitive sequencing methods, such as SiMSen-Seq, enable single-molecule detection by utilizing unique molecular identifiers. We demonstrate two applications of ultrasensitive sequencing in liquid biopsy. In the first approach, we analyzed ctDNA as a biomarker in patients diagnosed with GIST. In the second approach, we profiled the immune repertoire of γδ T-cells in healthy individuals.

Our specific conclusions are:

Paper I: We established quality controls to analyze the amount of amplifiable DNA, degree of cellular contamination, and PCR inhibition. We show theoretically that the amount of plasma, DNA, and number of somatic variants monitored affect the sensitivity of ctDNA analysis. We also show that assay performance depends on both amplicon length and target sequence.

Paper II: We developed a patient-specific SiMSen-Seq panel to analyze ctDNA in patients diagnosed with GIST. Data showed that ctDNA analysis using patient-specific SiMSen-Seq panels can be applied to monitoring treatment efficacy and identifying relapse early, especially among high-risk patients.

Paper III: We developed the first targeted ultrasensitive immune repertoire sequencing approach to profile γδ T-cell clonality at DNA level. We showed that our approach could be applied to both enriched and non-enriched material, providing accurate quantification and minimizing sequencing errors.

# Future Perspective

Liquid biopsy analyses and precision medicine are driving a paradigm shift in both cancer management and other diseases. Adoption will be driven by demonstrating utility in health care as well as more simplified and cost-efficient workflows. In parallel, more complex biomarkers will be discovered, potentially with both increased sensitivity and specificity.

## Emerging clinical adaptations

While the field of ctDNA analysis has been developing for years and the technology's potential has been proven multiple times in proof-of-concept studies [20], demonstrating real clinical utility is still a challenge that needs attention and future studies [200]. Few studies utilizing ctDNA as a biomarker have proven clinical benefit over current practices to detect cancer early [201]. Studies are either too small or lack appropriate controls or comparison between methods. The number of FDA-approved diagnostics based on ctDNA are still few, and approvals are limited to use when a tissue biopsy is not applicable and complemented with tissue biopsy if negative [133,134]. More extensive studies are costly, delaying clinical implementation. Similar to the development of new drugs, one cannot expect academic researchers or government-supported organizations alone to conduct the necessary trials to implement ctDNA analysis in clinical routine broadly. Interventional studies need to be run with arms comparing patients monitored with liquid biopsy to conventional methods. Examples of such clinical trials are NCT03748680 and NCT04501523, which test if an adjuvant treatment given to ctDNA-positive patients is beneficial to the standard of care in colon and breast cancer respectively [202,203]. These studies take time to design, recruit, and evaluate, so clinical implementation will always lag behind cutting-edge development (Figure 15). Guardant 360 and FoundationOne, both based on molecular barcoding and targeted sequencing, were FDA approved for clinical use in 2020, almost a decade after similar techniques were first used to detect ctDNA in cancer [80,204,205]. Global adaptation of these technologies will take a much longer time.

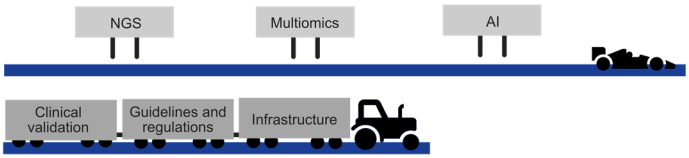


*Figure 15. Cutting-edge liquid biopsy technologies develop quickly. Clinical implementation takes time but is it the end what is contributing to patients' health.*

## Novel biomarkers and diagnostics

Liquid biopsy biomarkers and diagnostics are also becoming more complex. Today most common diagnostics evaluate a single biomarker using a single analyte, e.g., one protein, one mutation. Novel diagnostics can measure multiple biomarkers or report hundreds of analytes as a composite score. One such approach is CancerSEEK [206], which utilizes a combination of proteins and ctDNA and a novel machine-learning algorithm to detect nonmetastatic clinical detectable cancers with 70 % sensitivity, more than 99 % specificity, and the ability to predict the tumor location. When evaluated in a large screening-study enrolling more than 10 000 women with no prior history of cancer, the study almost doubled the number of patients identified by screening compared to the traditional program, maintaining a high specificity of 99.6 % [207]. These types of more complex tests present three challenges. First, it could increase the number of non-specific diagnoses that motivate investigation but do not lead to a cancer diagnosis. Second, a diagnostic may provide more information than the test was intended to generate, resulting in the need for additional investigations and negatively affect the quality of life of these patients. Third, more advanced algorithms using machine learning may generate a 'black box' issue, where it is difficult to understand why a test became positive or negative.

Besides ctDNA, analytes such as proteins, different RNA such as miRNA and lncRNA, as well as DNA methylation are promising biomarkers in next-generation liquid biopsies [208,209]. For example, one of the most promising biomarkers in pancreatic cancer is a multiparametric signature of 29 proteins [210]. Most emerging biomarker solutions have in common that they assess multiple markers, since individual biomarkers, such as a specific protein, miRNA, or methylation site, cannot provide a specific and sensitive answer on

its own. Detecting an array of proteins, miRNA or methylation sites result in increased specificity and sensitivity.

All strategies to detect cancer in a liquid biopsy are dependent on the amount of signal from tumor cells versus the amount of background from non-tumor cells. The biological processes of releasing the biomarker into the blood can be due to apoptosis, necrosis, cell growth, location, cell-to-cell communication, and several other biological processes affecting the signal-to-background relation. Figure 16 illustrates the pros and cons of using three types of analytes, including mutations in cfDNA, methylated sites in cfDNA, and proteins. Each analyte has a different signal-to-background ratio that is generated from tumor and non-tumor cells, respectively. Several factors influence this ratio. For example, cfDNA is mainly released into the bloodstream when cells die. In contrast, many proteins are predominantly secreted by active cells. Proteins and methylation markers usually have a high background from non-tumor cells, while mutations are more frequently unique to the tumor. These considerations are similar for all analytes and useful to consider when constructing a sensitive and specific test. Interestingly, prominent companies in liquid biopsy-based cancer screening, such as GRAIL, have moved away from mutations in favor of epigenetic-based approaches, anticipating that the high number of differentially methylated sites over somatic variants will provide increased sensitivity [211]. A potential disadvantage with this approach is that these epigenetic markers are primarily cell-of-origin patterns that are complicated to separate from other tissue damage [212].

A significant challenge associated with testing and screening can be the anxiety associated with the disease. More research is needed to understand the impact on quality of life when a patient receives a test result and how to communicate these results in the best possible way [213]. Importantly, anxiety is far from the only negative associated with false-positive and inconclusive results. A follow-up test may also result in physical harm, such as radiation exposure [214] or complications from invasive biopsy [215]. Decreasing these negative consequences would also enable more testing. Notably, liquid biopsy analysis plays an essential role in this emerging area as it needs to be as minimally invasive as possible.

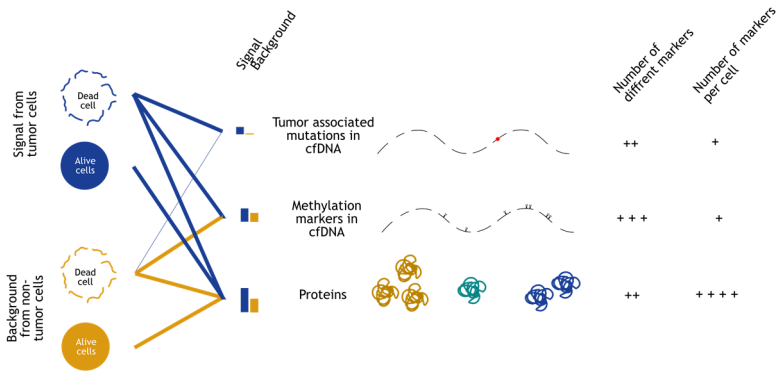


*Figure 16. Overview of analytes in liquid biopsies. Analytes can be released from tumor and non-tumor cells by several biological processes. Analytes from tumor cells are the biomarkers intended to be detected, while identical analytes from all other non-tumor cells generate the background. Each analyte has few or many cancer-associated biomarkers, each with a different signal-to-background ratio. The number of biomarkers released from a single cell is also different between analytes. For example, each cell contains only one genome but thousands of molecules for a given protein.*

Data availability on the performance of diagnostics tests based on real-world clinical data is another significant driver for developing new diagnostics. For therapeutic drugs, physicians are supposed to report adverse events associated with use. A similar system could be used for diagnostics to report false positives and negatives. Resources must also be provided to include exploratory biomarkers parallel to clinical use, enabling tighter integration between research and clinic. In these days of privacy concerns and legislation, it is essential to enable professional solutions with respect to privacy and integrity—solutions that allow us to share more data in the future, not less.

## Personalized biomarkers

Finally, the idea that a single biomarker could detect all cancer with high specificity is as likely to become reality as there being a single drug to cure all cancer. Much like personalized medicine describes when to use specific treatments, a similar term is lacking to describe a selection of highly specific and sensitive "personalized biomarkers." For example, not all women will benefit from early mammography. Before the age of 50, screening decreases the risk of mortality equivalent to the increased risk of a fifteen-hour bike ride without a helmet [216]. In Sweden, screening starts at the age of 40 [217]. However, for specific risk populations, such as women who carry genes for

hereditary cancers, screening should be performed at a younger age and a higher frequency. Such mammography could be seen as a personalized biomarker. Our GIST study is also an example of a personalized biomarker, as the test was adopted to follow a patient-specific tumor mutation. An extreme form of a personalized biomarker is phylogenetic ctDNA tracking, in which multiple personalized assays based on patient tumor biopsies are developed to track minimal residual disease [172]. All these personalized biomarkers strive to increase sensitivity, specificity, and clinical utility for a selected patient population and point to an exciting future for biomarkers and diagnostics.

# Acknowledgment

This thesis would not have been possible without an extraordinary group of colleagues, friends, and organizations passionate about science. I would therefore like to express my appreciation to:

My main supervisor **Anders Ståhlberg**. Through your support and the freedom you gave me, I found my place as a scientist and grew professionally every day.

My industrial co-supervisors **Outi Vaarala** and **Michael Hühn** at AstraZeneca. Without your entrepreneurship, this industrial-academic collaboration would not have happened. My academic co-supervisor **Ruth Palmer** for being an exceptional scientist and role model.

**Ståhlberg group** for welcoming me to your lab. **Stefan and Daniel,** without your help throughout, but particularly in the beginning, I would still be a fish out of water. **Melita** and **Tamara,** you were the best students I can imagine! Thank you for allowing me to practice my boss skills on you! **Helena, Maria, Tobias, Soheila, Emma, Parmida, Lisa, Manuel, Johanna, Pia, Peter, and former Ståhlberg lab members** for making work fun even when it is not. **Colleagues at Sahlgrenska Center for Cancer Research, in particular, members of Åman group, Landberg group, Dalin group** for collaborations, fikas, meetings, and seminars and creating a great place to work. **Friends on the fourth floor,** including **Junchi, Daniel, Anna, Elin, and Karin,** let the candy-jar always be emptied as quickly as it is refilled! **Colleagues at AstraZeneca** for fun parties and valuable industry insights. **Marta, Maria, Stefan, and Andreas at the endocrine and sarcoma surgery section at Sahlgrenska University hospital** for our collaboration in the GIST study. **Cardell group** for introducing me to the fascinating world of immune repertoire analysis.

**All the patients** who donated samples used in these studies, not for their own immediate benefit but for the patients who will follow their path. It is our privilege to work with these samples and our responsibility to use it to perform the best possible science.

AstraZeneca; Knut and Alice Wallenberg Foundation, Wallenberg Centre for Molecular and Translational Medicine at University of Gothenburg: Assar Gabrielssons Research Foundation; BioCARE National Strategic Research Program at University of Gothenburg; Johan Jansson Foundation for Cancer Research; Anna-Lisa och Bror Björnsson Stiftelse; Swedish Childhood Cancer Foundation; Swedish Cancer Society; Swedish Research Council; the Swedish state under the agreement between the Swedish government and the county councils, the ALF-agreement and the Sweden's Innovation Agency for their funding support.

**My friends and family. "Australien-gänget**." **My grandfather Evert** for great inspiration in pursuing the unknown. **My parents** for fostering curiosity and passion for science. **My dear Bea** for endless support and love.

# References


[1] "The Human Genome Project," *Genome.gov*. https://www.genome.gov/human-genome-project (accessed Feb. 24, 2021).

[2] R. M. Durbin *et al.*, "A map of human genome variation from population-scale sequencing," *Nature*, vol. 467, no. 7319, Art. no. 7319, Oct. 2010, doi: 10/fmk7rw.

[3] A. Y. Maslov and J. Vijg, "Genome instability, cancer and aging," *Biochim. Biophys. Acta BBA - Gen. Subj.*, vol. 1790, no. 10, pp. 963–969, Oct. 2009, doi: 10/dxxtgv.

[4] M. Leija-Salazar, C. Piette, and C. Proukakis, "Review: Somatic mutations in neurodegeneration," *Neuropathol. Appl. Neurobiol.*, vol. 44, no. 3, pp. 267–285, Apr. 2018, doi: 10/gfgp5c.

[5] B. Vogelstein, N. Papadopoulos, V. E. Velculescu, S. Zhou, L. A. Diaz, and K. W. Kinzler, "Cancer Genome Landscapes," *Science*, vol. 339, no. 6127, pp. 1546–1558, Mar. 2013, doi: 10/6rg.

[6] M. Meyerson, S. Gabriel, and G. Getz, "Advances in understanding cancer genomes through second-generation sequencing," *Nat. Rev. Genet.*, vol. 11, no. 10, pp. 685–696, Oct. 2010, doi: 10/bdx8w2.

[7] C. Bettegowda *et al.*, "Detection of Circulating Tumor DNA in Early- and Late-Stage Human Malignancies," *Sci. Transl. Med.*, vol. 6, no. 224, pp. 224ra24-224ra24, Feb. 2014, doi: 10.1126/scitranslmed.3007094.

[8] M. Ignatiadis, G. W. Sledge, and S. S. Jeffrey, "Liquid biopsy enters the clinic — implementation issues and future challenges," *Nat. Rev. Clin. Oncol.*, pp. 1–16, Jan. 2021, doi: 10/ghvg9w.

[9] K. Shyamala, H. C. Girish, and S. Murgod, "Risk of tumor cell seeding through biopsy and aspiration cytology," *J. Int. Soc. Prev. Community Dent.*, vol. 4, no. 1, pp. 5–11, 2014, doi: 10.4103/2231-0762.129446.

[10] J. H. Boyum *et al.*, "Incidence and Risk Factors for Adverse Events Related to Image-Guided Liver Biopsy," *Mayo Clin. Proc.*, vol. 91, no. 3, pp. 329–335, Mar. 2016, doi: 10.1016/j.mayocp.2015.11.015.

[11] O. Bozbıyık, Ş. Öztürk, M. Ünver, V. Erol, Ü. Bayol, and C. Aydın, "Reliability of fine needle aspiration biopsy in large thyroid nodules," *Turk. J. Surg.*, vol. 33, no. 1, pp. 10–13, Mar. 2017, doi: 10.5152/UCD.2017.3329.

[12] D. Andersson, M. Kubista, and A. Ståhlberg, "Liquid biopsy analysis in cancer diagnostics," *Mol. Aspects Med.*, vol. 72, p. 100839, Apr. 2020, doi: 10.1016/j.mam.2019.100839.

[13] M. Gerlinger *et al.*, “Intratumor heterogeneity and branched evolution revealed by multiregion sequencing,” *N. Engl. J. Med.*, vol. 366, no. 10, pp. 883–892, Mar. 2012, doi: 10.1056/NEJMoa1113205.
[14] M. Murtaza *et al.*, “Non-invasive analysis of acquired resistance to cancer therapy by sequencing of plasma DNA,” *Nature*, vol. 497, no. 7447, pp. 108–112, May 2013, doi: 10.1038/nature12065.
[15] A. J. Bronkhorst, V. Ungerer, and S. Holdenrieder, “The emerging role of cell-free DNA as a molecular marker for cancer management,” *Biomol. Detect. Quantif.*, vol. 17, p. 100087, Mar. 2019, doi: 10.1016/j.bdq.2019.100087.
[16] S. Jahr *et al.*, “DNA Fragments in the Blood Plasma of Cancer Patients: Quantitations and Evidence for Their Origin from Apoptotic and Necrotic Cells,” *Cancer Res.*, vol. 61, no. 4, pp. 1659–1665, Feb. 2001.
[17] S. Nagata, H. Nagase, K. Kawane, N. Mukae, and H. Fukuyama, “Degradation of chromosomal DNA during apoptosis,” *Cell Death Differ.*, vol. 10, no. 1, Art. no. 1, Jan. 2003, doi: 10.1038/sj.cdd.4401161.
[18] M. Stroun, J. Lyautey, C. Lederrey, A. Olson-Sand, and P. Anker, “About the possible origin and mechanism of circulating DNA: Apoptosis and active DNA release,” *Clin. Chim. Acta*, vol. 313, no. 1, pp. 139–142, Nov. 2001, doi: 10.1016/S0009-8981(01)00665-9.
[19] N. B. Y. Tsui *et al.*, “High Resolution Size Analysis of Fetal DNA in the Urine of Pregnant Women by Paired-End Massively Parallel Sequencing,” *PLoS ONE*, vol. 7, no. 10, Oct. 2012, doi: 10.1371/journal.pone.0048319.
[20] J. C. M. Wan *et al.*, “Liquid biopsies come of age: Towards implementation of circulating tumour DNA,” *Nat. Rev. Cancer*, vol. 17, no. 4, pp. 223–238, Apr. 2017, doi: 10.1038/nrc.2017.7.
[21] Y. W. L. Zheng *et al.*, “Nonhematopoietically Derived DNA Is Shorter than Hematopoietically Derived DNA in Plasma: A Transplantation Model,” *Clin. Chem.*, vol. 58, no. 3, pp. 549–558, Mar. 2012, doi: 10.1373/clinchem.2011.169318.
[22] J. Moss *et al.*, “Comprehensive human cell-type methylation atlas reveals origins of circulating cell-free DNA in health and disease,” *Nat. Commun.*, vol. 9, no. 1, Art. no. 1, Nov. 2018, doi: 10.1038/s41467-018-07466-6.
[23] F. Mouliere *et al.*, “Enhanced detection of circulating tumor DNA by fragment size analysis,” *Sci. Transl. Med.*, vol. 10, no. 466, Nov. 2018, doi: 10.1126/scitranslmed.aat4921.
[24] A. V. Cherepanova, S. N. Tamkovich, O. E. Bryzgunova, V. V. Vlassov, and P. P. Laktionov, “Deoxyribonuclease Activity and Circulating DNA Concentration in Blood Plasma of Patients with Prostate Tumors,” *Ann. N. Y. Acad. Sci.*, vol. 1137, no. 1, pp. 218–221, 2008, doi: https://doi.org/10.1196/annals.1448.016.

[25] V. J. Gauthier, L. N. Tyler, and M. Mannik, “Blood clearance kinetics and liver uptake of mononucleosomes in mice,” *J. Immunol. Baltim. Md 1950*, vol. 156, no. 3, pp. 1151–1156, Feb. 1996.

[26] T. M. Chused, A. D. Steinberg, and N. Talal, “The clearance and localization of nucleic acids by New Zealand and normal mice,” *Clin. Exp. Immunol.*, vol. 12, no. 4, pp. 465–476, Dec. 1972.

[27] I. Botezatu *et al.*, “Genetic Analysis of DNA Excreted in Urine: A New Approach for Detecting Specific Genomic DNA Sequences from Cells Dying in an Organism,” *Clin. Chem.*, vol. 46, no. 8, pp. 1078–1084, Aug. 2000, doi: 10.1093/clinchem/46.8.1078.

[28] A. Kustanovich, R. Schwartz, T. Peretz, and A. Grinshpun, “Life and death of circulating cell-free DNA,” *Cancer Biol. Ther.*, vol. 20, no. 8, pp. 1057–1067, Apr. 2019, doi: 10.1080/15384047.2019.1598759.

[29] J. Atamaniuk, C. Vidotto, H. Tschan, N. Bachl, K. M. Stuhlmeier, and M. M. Müller, “Increased concentrations of cell-free plasma DNA after exhaustive exercise,” *Clin. Chem.*, vol. 50, no. 9, pp. 1668–1670, Sep. 2004, doi: 10.1373/clinchem.2004.034553.

[30] A. Paunel-Görgülü *et al.*, “cfDNA correlates with endothelial damage after cardiac surgery with prolonged cardiopulmonary bypass and amplifies NETosis in an intracellular TLR9-independent manner,” *Sci. Rep.*, vol. 7, no. 1, p. 17421, Dec. 2017, doi: 10.1038/s41598-017-17561-1.

[31] G. J. Fournié, F. Martres, J. P. Pourrat, C. Alary, and M. Rumeau, “Plasma DNA as Cell Death Marker in Elderly Patients,” *Gerontology*, vol. 39, no. 4, pp. 215–221, 1993, doi: 10.1159/000213536.

[32] Y. M. D. Lo, T. H. Rainer, L. Y. S. Chan, N. M. Hjelm, and R. A. Cocks, “Plasma DNA as a Prognostic Marker in Trauma Patients,” *Clin. Chem.*, vol. 46, no. 3, pp. 319–323, Mar. 2000, doi: 10.1093/clinchem/46.3.319.

[33] N. Jiang and D. S. Pisetsky, “The effect of inflammation on the generation of plasma DNA from dead and dying cells in the peritoneum,” *J. Leukoc. Biol.*, vol. 77, no. 3, pp. 296–302, 2005, doi: https://doi.org/10.1189/jlb.0704411.

[34] M. Haghiac *et al.*, “Increased death of adipose cells, a path to release cell free DNA into systemic circulation of obese women,” *Obes. Silver Spring Md*, vol. 20, no. 11, pp. 2213–2219, Nov. 2012, doi: 10/f4fmp5.

[35] Y.-M. D. Lo, J. S. Wainscoat, M. D. G. Gillmer, P. Patel, M. Sampietro, and K. A. Fleming, “Prenatal sex determination by dna amplification from maternal peripheral blood,” *The Lancet*, vol. 334, no. 8676, pp. 1363–1365, Dec. 1989, doi: 10.1016/S0140-6736(89)91969-7.

[36] Y. M. Lo *et al.*, “Presence of fetal DNA in maternal plasma and serum,” *Lancet Lond. Engl.*, vol. 350, no. 9076, pp. 485–487, Aug. 1997, doi: 10.1016/S0140-6736(97)02174-0.

[37] H. Nawroz, W. Koch, P. Anker, M. Stroun, and D. Sidransky, “Microsatellite alterations in serum DNA of head and neck cancer

patients," *Nat. Med.*, vol. 2, no. 9, pp. 1035–1037, Sep. 1996, doi: 10.1038/nm0996-1035.
[38] J. B. de Kok *et al.*, "Detection of tumour DNA in serum of colorectal cancer patients," *Scand. J. Clin. Lab. Invest.*, vol. 57, no. 7, pp. 601–604, Nov. 1997, doi: 10.3109/00365519709055283.
[39] F. Mouliere, S. Messaoudi, D. Pang, A. Dritschilo, and A. R. Thierry, "Multi-marker analysis of circulating cell-free DNA toward personalized medicine for colorectal cancer.," *Mol. Oncol.*, vol. 8, no. 5, pp. 927–41, Jul. 2014, doi: 10.1016/j.molonc.2014.02.005.
[40] R. Mead, M. Duku, P. Bhandari, and I. A. Cree, "Circulating tumour markers can define patients with normal colons, benign polyps, and cancers," *Br. J. Cancer*, 2011, doi: 10/cj7zzv.
[41] A. Piovesan, M. C. Pelleri, F. Antonaros, P. Strippoli, M. Caracausi, and L. Vitale, "On the length, weight and GC content of the human genome," *BMC Res. Notes*, vol. 12, no. 1, p. 106, Feb. 2019, doi: 10/ghk7rv.
[42] P. Razavi *et al.*, "High-intensity sequencing reveals the sources of plasma circulating cell-free DNA variants," *Nat. Med.*, vol. 25, no. 12, pp. 1928–1937, Dec. 2019, doi: 10.1038/s41591-019-0652-7.
[43] S. Jaiswal *et al.*, "Age-Related Clonal Hematopoiesis Associated with Adverse Outcomes," *N. Engl. J. Med.*, vol. 371, no. 26, pp. 2488–2498, Dec. 2014, doi: 10/xdq.
[44] J. J. Salk *et al.*, "Ultra-Sensitive TP53 Sequencing for Cancer Detection Reveals Progressive Clonal Selection in Normal Tissue over a Century of Human Lifespan," *Cell Rep.*, vol. 28, no. 1, pp. 132-144.e3, Jul. 2019, doi: 10/ghxptp.
[45] J. H. Ginkel *et al.*, "Preanalytical blood sample workup for cell-free DNA analysis using Droplet Digital PCR for future molecular cancer diagnostics," *Cancer Med.*, vol. 6, no. 10, pp. 2297–2307, Oct. 2017, doi: 10.1002/cam4.1184.
[46] R. M. Trigg, L. J. Martinson, S. Parpart-Li, and J. A. Shaw, "Factors that influence quality and yield of circulating-free DNA: A systematic review of the methodology literature," *Heliyon*, vol. 4, no. 7, p. e00699, Jul. 2018, doi: 10/gd2f8s.
[47] S. R. Greytak *et al.*, "Harmonizing Cell-Free DNA Collection and Processing Practices through Evidence-Based Guidance," *Clin. Cancer Res.*, vol. 26, no. 13, pp. 3104–3109, Jul. 2020, doi: 10/gh482n.
[48] B. Risberg *et al.*, "Effects of Collection and Processing Procedures on Plasma Circulating Cell-Free DNA from Cancer Patients," *J. Mol. Diagn.*, vol. 20, no. 6, pp. 883–892, Nov. 2018, doi: 10/gfmm2b.
[49] K. A. Chan, S.-W. Yeung, W.-B. Lui, T. H. Rainer, and Y. D. Lo, "Effects of Preanalytical Factors on the Molecular Size of Cell-Free DNA in Blood," *Clin. Chem.*, vol. 51, no. 4, pp. 781–784, Apr. 2005, doi: 10/bh36pj.

[50] A. Sato *et al.*, “Investigation of appropriate pre-analytical procedure for circulating free DNA from liquid biopsy,” *Oncotarget*, vol. 9, no. 61, pp. 31904–31914, Aug. 2018, doi: 10.18632/oncotarget.25881.
[51] H. Markus *et al.*, “Evaluation of pre-analytical factors affecting plasma DNA analysis,” *Sci. Rep.*, vol. 8, no. 1, p. 7375, Dec. 2018, doi: 10.1038/s41598-018-25810-0.
[52] L. L. Lanier, J. Ruitenberg, R. L. Bolhuis, J. Borst, J. H. Phillips, and R. Testi, “Structural and serological heterogeneity of gamma/delta T cell antigen receptor expression in thymus and peripheral blood,” *Eur. J. Immunol.*, vol. 18, no. 12, pp. 1985–1992, Dec. 1988, doi: 10.1002/eji.1830181218.
[53] R. K. Saiki *et al.*, “Enzymatic amplification of beta-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia,” *Science*, vol. 230, no. 4732, pp. 1350–1354, Dec. 1985, doi: 10.1126/science.2999980.
[54] S. J. Scharf, G. T. Horn, and H. A. Erlich, “Direct cloning and sequence analysis of enzymatically amplified genomic sequences,” *Science*, vol. 233, no. 4768, pp. 1076–1078, Sep. 1986, doi: 10.1126/science.3461561.
[55] K. B. Mullis, “Process for amplifying nucleic acid sequences,” US4683202A, Jul. 28, 1987.
[56] T. B. Morrison, J. J. Weis, and C. T. Wittwer, “Quantification of low-copy transcripts by continuous SYBR Green I monitoring during amplification,” *BioTechniques*, vol. 24, no. 6, pp. 954–958, 960, 962, Jun. 1998.
[57] M. W. Pfaffl, “A new mathematical model for relative quantification in real-time RT-PCR,” *Nucleic Acids Res.*, vol. 29, no. 9, p. e45, May 2001, doi: 10.1093/nar/29.9.e45.
[58] B. Vogelstein and K. W. Kinzler, “Digital PCR.,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 96, no. 16, pp. 9236–9241, Aug. 1999, doi: 10.1073/PNAS.96.16.9236.
[59] B. J. Hindson *et al.*, “High-throughput droplet digital PCR system for absolute quantitation of DNA copy number.,” *Anal. Chem.*, vol. 83, no. 22, pp. 8604–10, Nov. 2011, doi: 10.1021/ac202028g.
[60] V. Rowlands, A. J. Rutkowski, E. Meuser, T. H. Carr, E. A. Harrington, and J. C. Barrett, “Optimisation of robust singleplex and multiplex droplet digital PCR assays for high confidence mutation detection in circulating tumour DNA,” *Sci. Rep.*, vol. 9, no. 1, Art. no. 1, Sep. 2019, doi: 10.1038/s41598-019-49043-x.
[61] A. Rohlin, J. Wernersson, Y. Engwall, L. Wiklund, J. Björk, and M. Nordling, “Parallel sequencing used in detection of mosaic mutations: comparison with four diagnostic DNA screening techniques,” *Hum. Mutat.*, vol. 30, no. 6, pp. 1012–1020, Jun. 2009, doi: 10.1002/humu.20980.

[62] C. Baer *et al.*, “Ultra-deep sequencing leads to earlier and more sensitive detection of the tyrosine kinase inhibitor resistance mutation T315I in chronic myeloid leukemia,” *Haematologica*, vol. 101, no. 7, pp. 830–838, Jul. 2016, doi: 10.3324/haematol.2016.145888.

[63] G. Chen, S. Mosier, C. D. Gocke, M.-T. Lin, and J. R. Eshleman, “Cytosine Deamination Is a Major Cause of Baseline Noise in Next-Generation Sequencing,” *Mol. Diagn. Ther.*, vol. 18, no. 5, pp. 587–593, Oct. 2014, doi: 10.1007/s40291-014-0115-2.

[64] J. J. Salk, M. W. Schmitt, and L. A. Loeb, “Enhancing the accuracy of next-generation sequencing for detecting rare and subclonal mutations,” *Nat. Rev. Genet.*, vol. 19, no. 5, Art. no. 5, May 2018, doi: 10.1038/nrg.2017.117.

[65] P. J. A. Cock, C. J. Fields, N. Goto, M. L. Heuer, and P. M. Rice, “The Sanger FASTQ file format for sequences with quality scores, and the Solexa/Illumina FASTQ variants,” *Nucleic Acids Res.*, vol. 38, no. 6, pp. 1767–1771, Apr. 2010, doi: 10.1093/nar/gkp1137.

[66] K. Cibulskis *et al.*, “Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples,” *Nat. Biotechnol.*, vol. 31, no. 3, Art. no. 3, Mar. 2013, doi: 10.1038/nbt.2514.

[67] D. C. Koboldt *et al.*, “VarScan 2: Somatic mutation and copy number alteration discovery in cancer by exome sequencing,” *Genome Res.*, vol. 22, no. 3, pp. 568–576, Mar. 2012, doi: 10.1101/gr.129684.111.

[68] H. Li, “Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM,” *ArXiv13033997 Q-Bio*, May 2013, Accessed: Jan. 07, 2021. [Online]. Available: http://arxiv.org/abs/1303.3997.

[69] Z. Wei, W. Wang, P. Hu, G. J. Lyon, and H. Hakonarson, “SNVer: a statistical tool for variant calling in analysis of pooled or individual next-generation sequencing data,” *Nucleic Acids Res.*, vol. 39, no. 19, pp. e132–e132, Oct. 2011, doi: 10.1093/nar/gkr599.

[70] A. Wilm *et al.*, “LoFreq: a sequence-quality aware, ultra-sensitive variant caller for uncovering cell-population heterogeneity from high-throughput sequencing datasets,” *Nucleic Acids Res.*, vol. 40, no. 22, pp. 11189–11201, Dec. 2012, doi: 10.1093/nar/gks918.

[71] M. Gerstung *et al.*, “Reliable detection of subclonal single-nucleotide variants in tumour cell populations,” *Nat. Commun.*, vol. 3, no. 1, p. 811, Jan. 2012, doi: 10.1038/ncomms1814.

[72] X. Ma *et al.*, “Analysis of error profiles in deep next-generation sequencing data,” *Genome Biol.*, vol. 20, no. 1, p. 50, Mar. 2019, doi: 10.1186/s13059-019-1659-6.

[73] B. Arbeithuber, K. D. Makova, and I. Tiemann-Boege, “Artifactual mutations resulting from DNA lesions limit detection levels in ultrasensitive sequencing applications,” *DNA Res. Int. J. Rapid Publ. Rep. Genes Genomes*, vol. 23, no. 6, pp. 547–559, Dec. 2016, doi: 10.1093/dnares/dsw038.

[74] L. Chen, P. Liu, T. C. Evans, and L. M. Ettwiller, “DNA damage is a pervasive cause of sequencing errors, directly confounding variant identification,” *Science*, vol. 355, no. 6326, pp. 752–756, Feb. 2017, doi: 10.1126/science.aai8690.

[75] M. Costello *et al.*, “Discovery and characterization of artifactual mutations in deep coverage targeted capture sequencing data due to oxidative DNA damage during sample preparation,” *Nucleic Acids Res.*, vol. 41, no. 6, pp. e67–e67, Apr. 2013, doi: 10.1093/nar/gks1443.

[76] H. Do and A. Dobrovic, “Sequence Artifacts in DNA from Formalin-Fixed Tissues: Causes and Strategies for Minimization,” *Clin. Chem.*, vol. 61, no. 1, pp. 64–71, Jan. 2015, doi: 10.1373/clinchem.2014.223040.

[77] O. Zagordi, R. Klein, M. Däumer, and N. Beerenwinkel, “Error correction of next-generation sequencing data and reliable estimation of HIV quasispecies,” *Nucleic Acids Res.*, vol. 38, no. 21, pp. 7400–7409, Nov. 2010, doi: 10.1093/nar/gkq655.

[78] M. W. Schmitt, E. J. Fox, and J. J. Salk, “Risks of double-counting in deep sequencing,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 111, no. 16, p. E1560, Apr. 2014, doi: 10.1073/pnas.1400941111.

[79] K. Best, T. Oakes, J. M. Heather, J. Shawe-Taylor, and B. Chain, “Computational analysis of stochastic heterogeneity in PCR amplification efficiency revealed by single molecule barcoding,” *Sci. Rep.*, vol. 5, no. 1, Art. no. 1, Oct. 2015, doi: 10.1038/srep14629.

[80] I. Kinde, J. Wu, N. Papadopoulos, K. W. Kinzler, and B. Vogelstein, “Detection and quantification of rare mutations with massively parallel sequencing.,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 108, no. 23, pp. 9530–9535, Jun. 2011, doi: 10.1073/pnas.1105422108.

[81] J. Hong and D. Gresham, “Incorporation of unique molecular identifiers in TruSeq adapters improves the accuracy of quantitative sequencing,” *BioTechniques*, vol. 63, no. 5, pp. 221–226, Nov. 2017, doi: 10.2144/000114608.

[82] M. W. Schmitt, S. R. Kennedy, J. J. Salk, E. J. Fox, J. B. Hiatt, and L. A. Loeb, “Detection of ultra-rare mutations by next-generation sequencing.,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 109, no. 36, pp. 14508–14513, Sep. 2012, doi: 10.1073/pnas.1208715109.

[83] D. I. Lou *et al.*, “High-throughput DNA sequencing errors are reduced by orders of magnitude using circle sequencing,” *Proc. Natl. Acad. Sci.*, vol. 110, no. 49, pp. 19872–19877, 2013, doi: 10.1073/pnas.1319590110.

[84] C. Li, K. R. Chng, E. J. H. Boey, A. H. Q. Ng, A. Wilm, and N. Nagarajan, “INC-Seq: accurate single molecule reads using nanopore sequencing,” *GigaScience*, vol. 5, no. 1, p. 34, Dec. 2016, doi: 10.1186/s13742-016-0140-7.

[85] M. T. Gregory *et al.*, “Targeted single molecule mutation detection with massively parallel sequencing,” *Nucleic Acids Res.*, vol. 44, no. 3, pp. e22--e22, Feb. 2016, doi: 10.1093/nar/gkv915.

[86] A. Ståhlberg, P. M. Krzyzanowski, J. B. Jackson, M. Egyud, L. Stein, and T. E. Godfrey, “Simple, multiplexed, PCR-based barcoding of DNA enables sensitive mutation detection in liquid biopsies using sequencing.,” *Nucleic Acids Res.*, vol. 44, no. 11, p. e105, Jun. 2016, doi: 10.1093/nar/gkw224.

[87] Z. LIU, J. J. LIU, G. Liu, and T. Chen, “Methods and compositions for reducing non-specific amplification products,” US10421993B2, Sep. 24, 2019.

[88] I. Ahmed, F. A. Tucci, A. Aflalo, K. G. C. Smith, and R. J. M. Bashford-Rogers, “Ultrasensitive amplicon barcoding for next-generation sequencing facilitating sequence error and amplification-bias correction,” *Sci. Rep.*, vol. 10, no. 1, Art. no. 1, Jun. 2020, doi: 10.1038/s41598-020-67290-1.

[89] E. Hodges *et al.*, “Genome-wide in situ exon capture for selective resequencing,” *Nat. Genet.*, vol. 39, no. 12, Art. no. 12, Dec. 2007, doi: 10.1038/ng.2007.42.

[90] A. Gnirke *et al.*, “Solution hybrid selection with ultra-long oligonucleotides for massively parallel targeted sequencing,” *Nat. Biotechnol.*, vol. 27, no. 2, Art. no. 2, Feb. 2009, doi: 10.1038/nbt.1523.

[91] R. Chen, H. Im, and M. Snyder, “Whole-Exome Enrichment with the Agilent SureSelect Human All Exon Platform,” *Cold Spring Harb. Protoc.*, vol. 2015, no. 7, p. pdb.prot083659, Jul. 2015, doi: 10.1101/pdb.prot083659.

[92] R. Chen, H. Im, and M. Snyder, “Whole-Exome Enrichment with the Illumina TruSeq Exome Enrichment Platform,” *Cold Spring Harb. Protoc.*, vol. 2015, no. 7, p. pdb.prot084863, Jul. 2015, doi: 10.1101/pdb.prot084863.

[93] D. Schenk, G. Song, Y. Ke, and Z. Wang, “Amplification of overlapping DNA amplicons in a single-tube multiplex PCR for targeted next-generation sequencing of BRCA1 and BRCA2,” *PLOS ONE*, vol. 12, no. 7, p. e0181062, Jul. 2017, doi: 10.1371/journal.pone.0181062.

[94] R. B. Lanman *et al.*, “Analytical and Clinical Validation of a Digital Sequencing Panel for Quantitative, Highly Accurate Evaluation of Cell-Free Circulating Tumor DNA,” *PLoS ONE*, vol. 10, no. 10, Oct. 2015, doi: 10.1371/journal.pone.0140712.

[95] A. Ståhlberg, P. M. Krzyzanowski, M. Egyud, S. Filges, L. Stein, and T. E. Godfrey, “Simple multiplexed PCR-based barcoding of DNA for ultrasensitive mutation detection by next-generation sequencing,” *Nat. Protoc.*, vol. 12, no. 4, pp. 664–682, 2017, doi: 10.1038/nprot.2017.006.

[96] M. Egyud *et al.*, "Detection of Circulating Tumor DNA in Plasma: A Potential Biomarker for Esophageal Adenocarcinoma.," *Ann. Thorac. Surg.*, vol. 108, no. 2, pp. 343–349, Aug. 2019, doi: 10/ghxqqc.
[97] S. Bjursten *et al.*, "Response to BRAF/MEK Inhibition in A598_T599insV BRAF Mutated Melanoma.," *Case Rep. Oncol.*, vol. 12, no. 3, pp. 872–879, Dec. 2019, doi: 10/ghxqp8.
[98] C. Vannas, S. Bjursten, S. Filges, H. Fagman, A. Ståhlberg, and M. Levin, "Dynamic ctDNA evaluation of a patient with BRAFV600E metastatic melanoma demonstrates the utility of ctDNA for disease monitoring and tumor clonality analysis," *Acta Oncol.*, pp. 1–5, Aug. 2020, doi: 10.1080/0284186X.2020.1802064.
[99] S. Herrmann *et al.*, "Detection of mutational patterns in cell-free DNA of colorectal cancer by custom amplicon sequencing.," *Mol. Oncol.*, vol. 13, no. 8, pp. 1669–1683, Aug. 2019, doi: 10/ghmkb4.
[100] T. Moser *et al.*, "On-treatment measurements of circulating tumor DNA during FOLFOX therapy in patients with colorectal cancer.," *NPJ Precis. Oncol.*, vol. 4, no. 1, p. 30, Nov. 2020, doi: 10/ghxqp4.
[101] M. Egyud *et al.*, "Plasma circulating tumor DNA as a potential tool for disease monitoring in head and neck cancer.," *Head Neck*, vol. 41, no. 5, pp. 1351–1358, May 2019, doi: 10/ghxqqg.
[102] S. Mallampati *et al.*, "Rational 'Error Elimination' Approach to Evaluating Molecular Barcoded Next-Generation Sequencing Data Identifies Low-Frequency Mutations in Hematologic Malignancies.," *J. Mol. Diagn. JMD*, vol. 21, no. 3, pp. 471–482, May 2019, doi: 10/gg7h6w.
[103] A. Nakamura *et al.*, "Detection of MYD88 L265P mutation by next-generation deep sequencing in peripheral blood mononuclear cells of Waldenström's macroglobulinemia and IgM monoclonal gammopathy of undetermined significance.," *PloS One*, vol. 14, no. 9, p. e0221941, 2019, doi: 10/ghxqp9.
[104] G. Pabst *et al.*, "TP53 mutated AML subclones exhibit engraftment in a humanized bone marrow ossicle mouse model.," *Ann. Hematol.*, vol. 99, no. 3, pp. 653–655, Mar. 2020, doi: 10/ghxqp7.
[105] S. Filges, E. Yamada, A. Ståhlberg, and T. E. Godfrey, "Impact of Polymerase Fidelity on Background Error Rates in Next-Generation Sequencing with Unique Molecular Identifiers/Barcodes.," *Sci. Rep.*, vol. 9, no. 1, p. 3503, Mar. 2019, doi: 10/ghxqqd.
[106] K. Elliott *et al.*, "Elevated pyrimidine dimer formation at distinct genomic bases underlies promoter mutation hotspots in UV-exposed cancers," *PLOS Genet.*, vol. 14, no. 12, p. e1007849, Dec. 2018, doi: 10/ghxsjf.
[107] A.-L. Boutigny, A. Barranger, C. De Boisséson, Y. Blanchard, and M. Rolland, "Targeted Next Generation Sequencing to study insert stability in genetically modified plants.," *Sci. Rep.*, vol. 9, no. 1, p. 2308, Feb. 2019, doi: 10/ghxqqf.

[108] H.-J. He *et al.*, “Multilaboratory Assessment of a New Reference Material for Quality Assurance of Cell-Free Tumor DNA Measurements.,” *J. Mol. Diagn. JMD*, vol. 21, no. 4, pp. 658–676, Jul. 2019, doi: 10/ggdcvk.
[109] F. Diehl *et al.*, “Detection and quantification of mutations in the plasma of patients with colorectal tumors,” *Proc. Natl. Acad. Sci.*, vol. 102, no. 45, pp. 16368–16373, Nov. 2005, doi: 10.1073/pnas.0507904102.
[110] A. M. Newman *et al.*, “An ultrasensitive method for quantitating circulating tumor DNA with broad patient coverage.,” *Nat. Med.*, vol. 20, no. 5, pp. 548–54, May 2014, doi: 10.1038/nm.3519.
[111] R. Xu *et al.*, “Circulating tumour DNA methylation markers for diagnosis and prognosis of hepatocellular carcinoma,” *Nat. Mater.*, vol. 16, no. 11, pp. 1155–1161, Nov. 2017, doi: 10.1038/nmat4997.
[112] S. Avanzini *et al.*, “A mathematical model of ctDNA shedding predicts tumor detection size,” *Sci. Adv.*, vol. 6, no. 50, p. eabc4308, Dec. 2020, doi: 10.1126/sciadv.abc4308.
[113] S. Srivastava *et al.*, “Cancer overdiagnosis: a biological challenge and clinical dilemma,” *Nat. Rev. Cancer*, vol. 19, no. 6, pp. 349–358, Jun. 2019, doi: 10.1038/s41568-019-0142-8.
[114] D. Chen *et al.*, “Liquid Biopsy Applications in the Clinic,” *Mol. Diagn. Ther.*, vol. 24, no. 2, pp. 125–132, Apr. 2020, doi: 10.1007/s40291-019-00444-8.
[115] S. M. Esagian *et al.*, “Comparison of liquid-based to tissue-based biopsy analysis by targeted next generation sequencing in advanced non-small cell lung cancer: a comprehensive systematic review,” *J. Cancer Res. Clin. Oncol.*, vol. 146, no. 8, pp. 2051–2066, Aug. 2020, doi: 10.1007/s00432-020-03267-x.
[116] T. Mok *et al.*, “Detection and Dynamic Changes of EGFR Mutations from Circulating Tumor DNA as a Predictor of Survival Outcomes in NSCLC Patients Treated with First-line Intercalated Erlotinib and Chemotherapy,” *Clin. Cancer Res.*, vol. 21, no. 14, pp. 3196–3203, Jul. 2015, doi: 10.1158/1078-0432.CCR-14-2594.
[117] O. A. Zill *et al.*, “The Landscape of Actionable Genomic Alterations in Cell-Free Circulating Tumor DNA from 21,807 Advanced Cancer Patients,” *Clin. Cancer Res.*, vol. 24, no. 15, pp. 3528–3538, Aug. 2018, doi: 10.1158/1078-0432.CCR-17-3837.
[118] E. Christensen *et al.*, “Early Detection of Metastatic Relapse and Monitoring of Therapeutic Efficacy by Ultra-Deep Sequencing of Plasma Cell-Free DNA in Patients With Urothelial Bladder Carcinoma,” *J. Clin. Oncol.*, vol. 37, no. 18, pp. 1547–1557, May 2019, doi: 10.1200/JCO.18.02052.
[119] Y. Wang *et al.*, “Prognostic Potential of Circulating Tumor DNA Measurement in Postoperative Surveillance of Nonmetastatic Colorectal Cancer,” *JAMA Oncol.*, May 2019, doi: 10.1001/jamaoncol.2019.0512.

[120] National Research Council (US), *Toward Precision Medicine*. National Academies Press (US), 2011.
[121] K. C. S. Oliveira *et al.*, "Current Perspectives on Circulating Tumor DNA, Precision Medicine, and Personalized Clinical Management of Cancer," *Mol. Cancer Res.*, vol. 18, no. 4, pp. 517–528, Apr. 2020, doi: 10.1158/1541-7786.MCR-19-0768.
[122] N. Hanna *et al.*, "Systemic Therapy for Stage IV Non–Small-Cell Lung Cancer: American Society of Clinical Oncology Clinical Practice Guideline Update," *J. Clin. Oncol.*, Aug. 2017, doi: 10.1200/JCO.2017.74.6065.
[123] T. S. Mok *et al.*, "Osimertinib or Platinum-Pemetrexed in EGFR T790M-Positive Lung Cancer," *N. Engl. J. Med.*, vol. 376, no. 7, pp. 629–640, Feb. 2017, doi: 10.1056/NEJMoa1612674.
[124] E. S. Gray *et al.*, "Circulating tumor DNA to monitor treatment response and detect acquired resistance in patients with metastatic melanoma," *Oncotarget*, vol. 6, no. 39, pp. 42008–42018, Dec. 2015, doi: 10.18632/oncotarget.5788.
[125] Z. Hélias-Rodzewicz *et al.*, "Variation of mutant allele frequency in NRAS Q61 mutated melanomas," *BMC Dermatol.*, vol. 17, no. 1, p. 9, Jul. 2017, doi: 10.1186/s12895-017-0061-x.
[126] A. A. Azad *et al.*, "Androgen Receptor Gene Aberrations in Circulating Cell-Free DNA: Biomarkers of Therapeutic Resistance in Castration-Resistant Prostate Cancer," *Clin. Cancer Res.*, vol. 21, no. 10, pp. 2315–2324, May 2015, doi: 10.1158/1078-0432.CCR-14-2666.
[127] S. Chandarlapaty *et al.*, "Prevalence of ESR1 Mutations in Cell-Free DNA and Outcomes in Metastatic Breast Cancer: A Secondary Analysis of the BOLERO-2 Clinical Trial," *JAMA Oncol.*, vol. 2, no. 10, pp. 1310–1315, Oct. 2016, doi: 10.1001/jamaoncol.2016.1279.
[128] F. André *et al.*, "Alpelisib for PIK3CA-Mutated, Hormone Receptor-Positive Advanced Breast Cancer," *N. Engl. J. Med.*, vol. 380, no. 20, pp. 1929–1940, May 2019, doi: 10.1056/NEJMoa1813904.
[129] A. R. Thierry *et al.*, "Clinical utility of circulating DNA analysis for rapid detection of actionable mutations to select metastatic colorectal patients for anti-EGFR treatment," *Ann. Oncol. Off. J. Eur. Soc. Med. Oncol.*, vol. 28, no. 9, pp. 2149–2159, Sep. 2017, doi: 10.1093/annonc/mdx330.
[130] U.S. Food and Drug Administration, "FDA approves liquid biopsy NGS companion diagnostic test for multiple cancers and biomarkers," Nov. 09, 2020. https://www.fda.gov/drugs/fda-approves-liquid-biopsy-ngs-companion-diagnostic-test-multiple-cancers-and-biomarkers (accessed Jan. 13, 2021).
[131] U.S. Food and Drug Administration, "FDA Approves First Liquid Biopsy Next-Generation Sequencing Companion Diagnostic Test," *FDA*, Aug. 07, 2020. https://www.fda.gov/news-events/press-

announcements/fda-approves-first-liquid-biopsy-next-generation-sequencing-companion-diagnostic-test (accessed Aug. 09, 2020).
[132] U.S. Food and Drug Administration, “cobas EGFR Mutation Test v2,” Nov. 03, 2018. https://www.fda.gov/drugs/resources-information-approved-drugs/cobas-egfr-mutation-test-v2 (accessed Jan. 13, 2021).
[133] U.S. Food and Drug Administration, “The therascreen PIK3CA RGQ PCR Kit - P190001 and P190004,” *FDA*, Dec. 20, 2019. https://www.fda.gov/medical-devices/recently-approved-devices/therascreen-pik3ca-rgq-pcr-kit-p190001-and-p190004 (accessed Jan. 13, 2021).
[134] G. D. Demetri *et al.*, “NCCN Task Force report: update on the management of patients with gastrointestinal stromal tumors,” *J. Natl. Compr. Cancer Netw. JNCCN*, vol. 8 Suppl 2, pp. S1-41; quiz S42-44, Apr. 2010, doi: 10.6004/jnccn.2010.0116.
[135] L. G. Kindblom, H. E. Remotti, F. Aldenborg, and J. M. Meis-Kindblom, “Gastrointestinal pacemaker cell tumor (GIPACT): gastrointestinal stromal tumors show phenotypic characteristics of the interstitial cells of Cajal,” *Am. J. Pathol.*, vol. 152, no. 5, pp. 1259–1269, May 1998.
[136] M. Miettinen and J. Lasota, Gastrointestinal stromal tumors - Definition, clinical, histological, immunohistochemical, and molecular genetic features and differential diagnosis. 2001.
[137] B. Nilsson *et al.*, “Gastrointestinal stromal tumors: The incidence, prevalence, clinical course, and prognostication in the preimatinib mesylate era - A population-based study in western Sweden,” *Cancer*, vol. 103, no. 4, pp. 821–829, Feb. 2005, doi: 10.1002/cncr.20862.
[138] G. D. Demetri *et al.*, “Efficacy and safety of sunitinib in patients with advanced gastrointestinal stromal tumour after failure of imatinib: a randomised controlled trial,” *Lancet Lond. Engl.*, vol. 368, no. 9544, pp. 1329–1338, Oct. 2006, doi: 10.1016/S0140-6736(06)69446-4.
[139] G. D. Demetri *et al.*, “Efficacy and Safety of Imatinib Mesylate in Advanced Gastrointestinal Stromal Tumors,” *http://dx.doi.org/10.1056/NEJMoa020461*, Oct. 2009, doi: 10.1056/nejmoa020461.
[140] T. Nishida, J. Y. Blay, S. Hirota, Y. Kitagawa, and Y. K. Kang, “The standard diagnosis, treatment, and follow-up of gastrointestinal stromal tumors based on guidelines,” 2016, doi: 10.1007/s10120-015-0526-8.
[141] M. Iwatsuki *et al.*, “Neoadjuvant and adjuvant therapy for gastrointestinal stromal tumors,” *Ann. Gastroenterol. Surg.*, vol. 3, no. 1, pp. 43–49, Sep. 2018, doi: 10.1002/ags3.12211.
[142] H. Joensuu *et al.*, “Risk of recurrence of gastrointestinal stromal tumour after surgery: an analysis of pooled population-based cohorts,” *Lancet Oncol.*, vol. 13, no. 3, pp. 265–274, Mar. 2012, doi: 10.1016/S1470-2045(11)70299-6.

[143] J. Maier *et al.*, “Detection of Mutant Free Circulating Tumor DNA in the Plasma of Patients with Gastrointestinal Stromal Tumor Harboring Activating Mutations of CKIT or PDGFRA,” *Clin. Cancer Res.*, vol. 19, no. 17, pp. 4854–4867, Sep. 2013, doi: 10.1158/1078-0432.CCR-13-0765.
[144] H. M. Namløs *et al.*, “Noninvasive Detection of ctDNA Reveals Intratumor Heterogeneity and Is Associated with Tumor Burden in Gastrointestinal Stromal Tumor,” *Mol. Cancer Ther.*, vol. 17, no. 11, pp. 2473–2480, Nov. 2018, doi: 10.1158/1535-7163.MCT-18-0174.
[145] S. Jilg *et al.*, “Circulating cKIT and PDGFRA DNA indicates disease activity in Gastrointestinal Stromal Tumor (GIST),” *Int. J. Cancer*, vol. 145, no. 8, pp. 2292–2303, 2019, doi: 10.1002/ijc.32282.
[146] C. Serrano *et al.*, “Clinical value of next generation sequencing of plasma cell-free DNA in gastrointestinal stromal tumors,” *BMC Cancer*, vol. 20, no. 1, p. 99, Dec. 2020, doi: 10.1186/s12885-020-6597-x.
[147] N. K. Jerne, “The somatic generation of immune recognition,” *Eur. J. Immunol.*, vol. 1, no. 1, pp. 1–9, 1971, doi: https://doi.org/10.1002/eji.1830010102.
[148] J. J. M. van Dongen *et al.*, “Design and standardization of PCR primers and protocols for detection of clonal immunoglobulin and T-cell receptor gene recombinations in suspect lymphoproliferations: Report of the BIOMED-2 Concerted Action BMH4-CT98-3936,” *Leukemia*, vol. 17, no. 12, Art. no. 12, Dec. 2003, doi: 10.1038/sj.leu.2403202.
[149] L. Schirmer, V. Rothhammer, B. Hemmer, and T. Korn, “Enriched CD161high CCR6+ γδ T Cells in the Cerebrospinal Fluid of Patients With Multiple Sclerosis,” *JAMA Neurol.*, vol. 70, no. 3, pp. 345–351, Mar. 2013, doi: 10.1001/2013.jamaneurol.409.
[150] G. Maimaitijiang *et al.*, “Association of Decreased Percentage of Vδ2+Vγ9+ γδ T Cells With Disease Severity in Multiple Sclerosis,” *Front. Immunol.*, vol. 9, 2018, doi: 10.3389/fimmu.2018.00748.
[151] D. G. Schatz and Y. Ji, “Recombination centres and the orchestration of V(D)J recombination,” *Nat. Rev. Immunol.*, vol. 11, no. 4, Art. no. 4, Apr. 2011, doi: 10.1038/nri2941.
[152] D. Baltimore, “Is terminal deoxynucleotidyl transferase a somatic mutagen in lymphocytes?,” *Nature*, vol. 248, no. 5447, Art. no. 5447, Mar. 1974, doi: 10.1038/248409a0.
[153] J. D. Galson, A. J. Pollard, J. Trück, and D. F. Kelly, “Studying the antibody repertoire after vaccination: practical applications,” *Trends Immunol.*, vol. 35, no. 7, pp. 319–331, Jul. 2014, doi: 10.1016/j.it.2014.04.005.
[154] R. J. M. Bashford-Rogers, K. G. C. Smith, and D. C. Thomas, “Antibody repertoire analysis in polygenic autoimmune diseases,” *Immunology*, vol. 155, no. 1, pp. 3–17, Sep. 2018, doi: 10.1111/imm.12927.

[155] Y. Zhuang, C. Zhang, Q. Wu, J. Zhang, Z. Ye, and Q. Qian, "Application of immune repertoire sequencing in cancer immunotherapy," *Int. Immunopharmacol.*, vol. 74, p. 105688, Sep. 2019, doi: 10.1016/j.intimp.2019.105688.
[156] S. D. Boyd *et al.*, "Measurement and Clinical Monitoring of Human Lymphocyte Clonality by Massively Parallel V-D-J Pyrosequencing," *Sci. Transl. Med.*, vol. 1, no. 12, pp. 12ra23-12ra23, Dec. 2009, doi: 10.1126/scitranslmed.3000540.
[157] W. Zhang *et al.*, "Characterization of the B Cell Receptor Repertoire in the Intestinal Mucosa and of Tumor-Infiltrating Lymphocytes in Colorectal Adenoma and Carcinoma," *J. Immunol. Baltim. Md 1950*, vol. 198, no. 9, pp. 3719–3728, May 2017, doi: 10.4049/jimmunol.1602039.
[158] L. Robert *et al.*, "CTLA4 Blockade Broadens the Peripheral T-Cell Receptor Repertoire," *Clin. Cancer Res.*, vol. 20, no. 9, pp. 2424–2432, May 2014, doi: 10.1158/1078-0432.CCR-13-2648.
[159] S. K. Subudhi *et al.*, "Clonal expansion of CD8 T cells in the systemic circulation precedes development of ipilimumab-induced toxicities," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 113, no. 42, pp. 11919–11924, Oct. 2016, doi: 10.1073/pnas.1611421113.
[160] O. of the Commissioner, "FDA authorizes first next generation sequencing-based test to detect very low levels of remaining cancer cells in patients with acute lymphoblastic leukemia or multiple myeloma," *FDA*, Mar. 24, 2020. https://www.fda.gov/news-events/press-announcements/fda-authorizes-first-next-generation-sequencing-based-test-detect-very-low-levels-remaining-cancer (accessed Jan. 14, 2021).
[161] M. Shugay *et al.*, "Towards error-free profiling of immune repertoires," *Nat. Methods*, vol. 11, no. 6, pp. 653–655, Jun. 2014, doi: 10.1038/nmeth.2960.
[162] M. A. Turchaninova *et al.*, "High-quality full-length immunoglobulin profiling with unique molecular barcoding," *Nat. Protoc.*, vol. 11, no. 9, Art. no. 9, Sep. 2016, doi: 10/f83ddk.
[163] L. He *et al.*, "Toward a more accurate view of human B-cell repertoire by next-generation sequencing, unbiased repertoire capture and single-molecule barcoding," *Sci. Rep.*, vol. 4, p. 6778, Oct. 2014, doi: 10.1038/srep06778.
[164] M. Bengtsson, A. Ståhlberg, P. Rorsman, and M. Kubista, "Gene expression profiling in single cells from the pancreatic islets of Langerhans reveals lognormal distribution of mRNA levels," *Genome Res.*, vol. 15, no. 10, pp. 1388–1392, Oct. 2005, doi: 10.1101/gr.3820805.
[165] A. Raj, C. S. Peskin, D. Tranchina, D. Y. Vargas, and S. Tyagi, "Stochastic mRNA synthesis in mammalian cells," *PLoS Biol.*, vol. 4, no. 10, p. e309, Oct. 2006, doi: 10.1371/journal.pbio.0040309.

[166] K. S. Lundberg, D. D. Shoemaker, M. W. Adams, J. M. Short, J. A. Sorge, and E. J. Mathur, "High-fidelity amplification using a thermostable DNA polymerase isolated from Pyrococcus furiosus," *Gene*, vol. 108, no. 1, pp. 1–6, Dec. 1991, doi: 10.1016/0378-1119(91)90480-y.
[167] B. Arezi and H. H. Hogrefe, "Escherichia coli DNA polymerase III epsilon subunit increases Moloney murine leukemia virus reverse transcriptase fidelity and accuracy of RT-PCR procedures," *Anal. Biochem.*, vol. 360, no. 1, pp. 84–91, Jan. 2007, doi: 10.1016/j.ab.2006.10.009.
[168] A. Ståhlberg, J. Håkansson, X. Xian, H. Semb, and M. Kubista, "Properties of the reverse transcription reaction in mRNA quantification," *Clin. Chem.*, vol. 50, no. 3, pp. 509–515, Mar. 2004, doi: 10.1373/clinchem.2003.026161.
[169] X. Liu *et al.*, "Systematic Comparative Evaluation of Methods for Investigating the TCRβ Repertoire," *PLOS ONE*, vol. 11, no. 3, p. e0152464, Mar. 2016, doi: 10/ghs4k9.
[170] P. Chovanec *et al.*, "Unbiased quantification of immunoglobulin diversity at the DNA level with VDJ-seq," *Nat. Protoc.*, vol. 13, no. 6, Art. no. 6, Jun. 2018, doi: 10/gdhq3x.
[171] C. Abbosh *et al.*, "Phylogenetic ctDNA analysis depicts early-stage lung cancer evolution," *Nature*, vol. 545, no. 7655, Art. no. 7655, May 2017, doi: 10/f9765m.
[172] E. Delsing Malmberg *et al.*, "Accurate and Sensitive Analysis of Minimal Residual Disease in Acute Myeloid Leukemia Using Deep Sequencing of Single Nucleotide Variations," *J. Mol. Diagn.*, vol. 21, no. 1, pp. 149–162, Jan. 2019, doi: 10/ghvj8n.
[173] G. Johansson, "Considerations and quality controls when analyzing cell-free tumor DNA," *Biomol. Detect. Quantif.*, p. 10, 2019.
[174] S. Cristiano *et al.*, "Genome-wide cell-free DNA fragmentation in patients with cancer," *Nature*, vol. 570, no. 7761, Art. no. 7761, Jun. 2019, doi: 10/ggrw24.
[175] T. D. Andrews, Y. Jeelall, D. Talaulikar, C. C. Goodnow, and M. A. Field, "DeepSNVMiner: a sequence analysis tool to detect emergent, rare mutations in subsets of cell populations," *PeerJ*, vol. 4, p. e2074, 2016, doi: 10/ghvkrx.
[176] M. Shugay *et al.*, "MAGERI: Computational pipeline for molecular-barcoded targeted resequencing," *PLoS Comput. Biol.*, vol. 13, no. 5, p. e1005480, May 2017, doi: 10/f96jz4.
[177] C. Xu *et al.*, "smCounter2: an accurate low-frequency variant caller for targeted sequencing data with unique molecular identifiers," *Bioinforma. Oxf. Engl.*, vol. 35, no. 8, pp. 1299–1309, Apr. 2019, doi: 10/gd7j28.
[178] *fulcrumgenomics/fgbio*. Fulcrum Genomics, 2021.

[179] Z. Lai *et al.*, "VarDict: a novel and versatile variant caller for next-generation sequencing in cancer research," *Nucleic Acids Res.*, vol. 44, no. 11, p. e108, Jun. 2016, doi: 10/f8v6qz.
[180] G. D. Demetri *et al.*, "Efficacy and safety of regorafenib for advanced gastrointestinal stromal tumours after failure of imatinib and sunitinib (GRID): an international, multicentre, randomised, placebo-controlled, phase 3 trial," *Lancet Lond. Engl.*, vol. 381, no. 9863, pp. 295–302, Jan. 2013, doi: 10/f2fsb4.
[181] G. Z. Li and C. P. Raut, "Targeted therapy and personalized medicine in gastrointestinal stromal tumors: drug resistance, mechanisms, and treatment strategies," *OncoTargets Ther.*, vol. 12, pp. 5123–5133, Jul. 2019, doi: 10/ghvp94.
[182] E. K. Evans *et al.*, "Abstract 791: BLU-285, the first selective inhibitor of PDGFRα D842V and KIT Exon 17 mutants," *Cancer Res.*, vol. 75, no. 15 Supplement, pp. 791–791, Aug. 2015, doi: 10/ghvqbz.
[183] G. Umapathy, P. Mendoza-Garcia, B. Hallberg, and R. H. Palmer, "Targeting anaplastic lymphoma kinase in neuroblastoma," *Apmis*, vol. 127, no. 5, pp. 288–302, May 2019, doi: 10/gf8xhh.
[184] A. T. Shaw *et al.*, "Resensitization to Crizotinib by the Lorlatinib ALK Resistance Mutation L1198F," *N. Engl. J. Med.*, vol. 374, no. 1, pp. 54–61, Jan. 2016, doi: 10/f75j4t.
[185] G. Ravegnini *et al.*, "Mechanisms of resistance to a PI3K inhibitor in gastrointestinal stromal tumors: an omic approach to identify novel druggable targets," *Cancer Manag. Res.*, vol. 11, pp. 6229–6244, Jul. 2019, doi: 10/ghvqh3.
[186] D. Hanahan and R. A. Weinberg, "Hallmarks of Cancer: The Next Generation," *Cell*, vol. 144, no. 5, pp. 646–674, Mar. 2011, doi: 10/b6snj3.
[187] G. Johansson *et al.*, "Ultrasensitive DNA Immune Repertoire Sequencing Using Unique Molecular Identifiers," *Clin. Chem.*, vol. 66, no. 9, pp. 1228–1237, Sep. 2020, doi: 10.1093/clinchem/hvaa159.
[188] M. S. Davey *et al.*, "Clonal selection in the human Vδ1 T cell repertoire indicates γδ TCR-dependent adaptive immune surveillance," *Nat. Commun.*, vol. 8, p. 14760, Mar. 2017, doi: 10.1038/ncomms14760.
[189] A. W. Langerak *et al.*, "EuroClonality/BIOMED-2 guidelines for interpretation and reporting of Ig/TCR clonality testing in suspected lymphoproliferations," *Leukemia*, vol. 26, no. 10, Art. no. 10, Oct. 2012, doi: 10.1038/leu.2012.246.
[190] B. Howie *et al.*, "High-throughput pairing of T cell receptor α and β sequences," *Sci. Transl. Med.*, vol. 7, no. 301, pp. 301ra131-301ra131, Aug. 2015, doi: 10/ghwj4k.
[191] M. A. Turchaninova *et al.*, "Pairing of T-cell receptor chains via emulsion PCR," *Eur. J. Immunol.*, vol. 43, no. 9, pp. 2507–2515, 2013, doi: 10/f2d496.

[192] N. Tickotsky, T. Sagiv, J. Prilusky, E. Shifrut, and N. Friedman, "McPAS-TCR: a manually curated catalogue of pathology-associated T cell receptor sequences," *Bioinforma. Oxf. Engl.*, vol. 33, no. 18, pp. 2924–2929, Sep. 2017, doi: 10/ghwh53.
[193] M. Shugay *et al.*, "VDJdb: a curated database of T-cell receptor sequences with known antigen specificity," *Nucleic Acids Res.*, vol. 46, no. D1, pp. D419–D427, Jan. 2018, doi: 10/gfc6xw.
[194] W. Zhang *et al.*, "PIRD: Pan Immune Repertoire Database," *Bioinformatics*, vol. 36, no. 3, pp. 897–903, Feb. 2020, doi: 10/ghwh58.
[195] P. Dash *et al.*, "Quantifiable predictive features define epitope-specific T cell receptor repertoires," *Nature*, vol. 547, no. 7661, Art. no. 7661, Jul. 2017, doi: 10/gbh5nm.
[196] M. Widrich *et al.*, "Modern Hopfield Networks and Attention for Immune Repertoire Classification," *bioRxiv*, p. 2020.04.12.038158, Aug. 2020, doi: 10/ggxv5x.
[197] A. F. Herrera and P. Armand, "Minimal Residual Disease Assessment in Lymphoma: Methods and Applications," *J. Clin. Oncol.*, vol. 35, no. 34, pp. 3877–3887, Dec. 2017, doi: 10/ghwndq.
[198] X. Liu and J. Wu, "History, applications, and challenges of immune repertoire research," *Cell Biol. Toxicol.*, vol. 34, no. 6, pp. 441–457, 2018, doi: 10.1007/s10565-018-9426-0.
[199] M. Ignatiadis, M. Lee, and S. S. Jeffrey, "Circulating Tumor Cells and Circulating Tumor DNA: Challenges and Opportunities on the Path to Clinical Utility," *Clin. Cancer Res.*, vol. 21, no. 21, pp. 4786–4800, Nov. 2015, doi: 10/f7xpcr.
[200] I. A. Cree *et al.*, "The evidence base for circulating tumour DNA blood-based biomarkers for the early detection of cancer: a systematic mapping review," *BMC Cancer*, vol. 17, Oct. 2017, doi: 10/gchfjz.
[201] K.-L. G. Spindler, "Implementing Non-invasive Circulating Tumor DNA Analysis to Optimize the Operative and Postoperative Treatment for Patients With Colorectal Cancer," clinicaltrials.gov, Clinical trial registration NCT03748680, Dec. 2020. Accessed: Jan. 28, 2021. [Online]. Available: https://clinicaltrials.gov/ct2/show/NCT03748680.
[202] L. Qiang, "A Prospective, Phase II Trial Using Circulating Tumor DNA to Initiate Post-operation Boost Therapy After Neoadjuvant Chemotherapy in Triple-negative Breast Cancer," clinicaltrials.gov, Clinical trial registration NCT04501523, Aug. 2020. Accessed: Feb. 16, 2021. [Online]. Available: https://clinicaltrials.gov/ct2/show/NCT04501523.
[203] C. H. D. Metz, M. Scheulen, N. Bornfeld, D. Lohmann, and M. Zeschnigk, "Ultradeep sequencing detects GNAQ and GNA11 mutations in cell-free DNA from plasma of patients with uveal melanoma," *Cancer Med.*, vol. 2, no. 2, pp. 208–215, 2013, doi: 10/ghwvz9.

[204] A. Talasaz *et al.*, "Use of the GUARDANT360 noninvasive tumor sequencing assay on 300 patients across colorectal, melanoma, lung, breast, and prostate cancers and its clinical utility.," *J. Clin. Oncol.*, vol. 32, no. 15_suppl, pp. e22041–e22041, May 2014, doi: 10/ghwv4p.
[205] J. D. Cohen *et al.*, "Detection and localization of surgically resectable cancers with a multi-analyte blood test.," *Science*, vol. 359, no. 6378, pp. 926–930, Feb. 2018, doi: 10.1126/science.aar3247.
[206] A. M. Lennon *et al.*, "Feasibility of blood testing combined with PET-CT to screen for cancer and guide intervention," *Science*, vol. 369, no. 6499, p. eabb9601, Jul. 2020, doi: 10/gg5mdh.
[207] D. N. P. Oliveira *et al.*, "Diagnostic plasma miRNA-profiles for ovarian cancer in patients with pelvic mass," *PLOS ONE*, vol. 14, no. 11, p. e0225249, Nov. 2019, doi: 10/ghxgng.
[208] C. Koelsche *et al.*, "Sarcoma classification by DNA methylation profiling," *Nat. Commun.*, vol. 12, no. 1, Art. no. 1, Jan. 2021, doi: 10/ghxf7s.
[209] L. D. Mellby *et al.*, "Serum Biomarker Signature-Based Liquid Biopsy for Diagnosis of Early-Stage Pancreatic Cancer," *J. Clin. Oncol. Off. J. Am. Soc. Clin. Oncol.*, vol. 36, no. 28, pp. 2887–2894, Oct. 2018, doi: 10/gfqzhc.
[210] K. Lasseter *et al.*, "Plasma cell-free DNA variant analysis compared with methylated DNA analysis in renal cell carcinoma," *Genet. Med.*, vol. 22, no. 8, Art. no. 8, Aug. 2020, doi: 10/gg3fkz.
[211] K. A. Hoadley *et al.*, "Cell-of-Origin Patterns Dominate the Molecular Classification of 10,000 Tumors from 33 Types of Cancer," *Cell*, vol. 173, no. 2, pp. 291-304.e6, Apr. 2018, doi: 10/gc7r87.
[212] H. D. Nelson, M. Pappas, A. Cantor, E. Haney, and R. Holmes, "Risk Assessment, Genetic Counseling, and Genetic Testing for BRCA-Related Cancer in Women: Updated Evidence Report and Systematic Review for the US Preventive Services Task Force," *JAMA*, vol. 322, no. 7, pp. 666–685, Aug. 2019, doi: 10/gh3mp6.
[213] J. M. Meulepas *et al.*, "Radiation Exposure From Pediatric CT Scans and Subsequent Cancer Risk in the Netherlands," *J. Natl. Cancer Inst.*, vol. 111, no. 3, pp. 256–263, Mar. 2019, doi: 10/gh3mp7.
[214] M. Borghesi *et al.*, "Complications After Systematic, Random, and Image-guided Prostate Biopsy," *Eur. Urol.*, vol. 71, no. 3, pp. 353–365, Mar. 2017, doi: 10/f9tqm2.
[215] "Mammograms Before 50 A Waste Of Time." http://www.scienceagogo.com/news/19980912094305data_trunc_sys.shtml (accessed Feb. 09, 2021).
[216] Regionala cancercentrum i samverkan, "Nationellt vårdprogram bröstcancer," Oct. 2020.