Advanced Drug Delivery Reviews 105 (2016) 3–19

Contents lists available at ScienceDirect

Advanced Drug Delivery Reviews

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

Diagnostics based on nucleic acid sequence variant profiling: PCR, hybridization, and NGS approaches☆

Dmitriy Khodakov [a,1], Chunyan Wang [a,1], David Yu Zhang [a,b,*]

[a] Department of Bioengineering, Rice University, Houston, TX, United States
[b] Department of Systems, Synthetic, and Physical Biology, Rice University, Houston, TX, United States

ARTICLE INFO

Article history:
Received 4 November 2015
Received in revised form 21 March 2016
Accepted 6 April 2016
Available online 16 April 2016

Keywords:
Molecular diagnostics
Nucleic acid tests
PCR
Hybridization
Next-generation sequencing

ABSTRACT

Nucleic acid sequence variations have been implicated in many diseases, and reliable detection and quantitation of DNA/RNA biomarkers can inform effective therapeutic action, enabling precision medicine. Nucleic acid analysis technologies being translated into the clinic can broadly be classified into hybridization, PCR, and sequencing, as well as their combinations. Here we review the molecular mechanisms of popular commercial assays, and their progress in translation into in vitro diagnostics.

Contents

☆ This review is part of the *Advanced Drug Delivery Reviews* theme issue on "Synthetic Biology: Innovative approaches for pharmaceutics and drug delivery".

* Corresponding author.
E-mail address: dyz1@rice.edu (D.Y. Zhang).

[1] Authors contributed equally.

http://dx.doi.org/10.1016/j.addr.2016.04.005

1. Introduction

DNA sequence variations are frequent among humans; by some estimates, any given pair of unrelated human genomes will differ by 1 nucleotide every 300 nucleotides [1,2]. The vast majority of these variations are likely to have small to no effect on phenotype because the variations are within introns or are silent mutations that do not change the translated amino acid sequence. Nevertheless, a large number of hereditary diseases are known to be caused by sequence variations in single genes [3,4] and molecular studies of cancer have highlighted the role of driver mutations in the growth and metastasis of tumors [5–7]. Within pathogen DNA as well, sequence variations have led to differences in impact on human health; antibiotics resistance is an emerging worldwide healthcare problem [8,9].

Many technologies for detection and quantitation of sequence variations have been developed for genomics and disease research. These technologies generally can be grouped into three approaches and their combinations: polymerase chain reaction (PCR), hybridization, and next-generation sequencing (NGS). Each approach has distinct technical and operational advantages and disadvantages, the comparison of which is the main focus of this review (Fig. 1). As of this writing a large number of PCR and hybridization assays have been cleared or approved by the United State Food and Drug Administration (FDA) as in vitro diagnostics (IVDs) [10,11,12]. NGS assays, in contrast, are still nascent in the realm of clinical diagnostics, as Illumina and Ion Torrent obtained FDA clearance only in 2014 [13]. Several companies, most notably Genomic Health [12], provide clinical NGS assays as laboratory developed tests (LDTs) rather than in vitro diagnostics (IVDs), but LDTs have also recently come under the scrutiny of the FDA [14,15]. See Box 1 for an explanation of LDTs and IVDs.

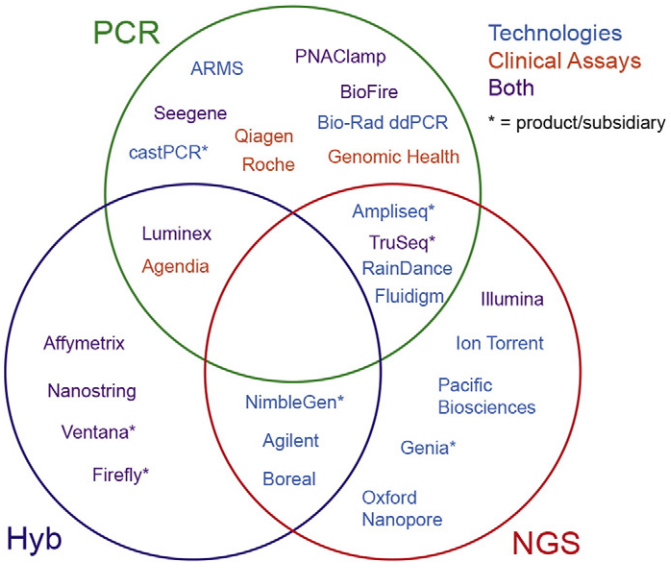

Fig. 1. Overview of technologies used for detection and profiling of nucleic acid sequence variations. The three broad approaches are PCR, hybridization, and NGS, but there is significant overlap between the three, and many technologies use a combination.

DNA sequence variation detection is valuable at many stages of a disease, and provides different clinical value at each stage; see Box 2 for a summary of subtypes of diagnostic tests by patient group and actionability. At a technical level, sequence variation detection must meet different levels of performance for the different applications. Detection of autosomal dominant germ line mutation for risk assessment, for example, requires only that 50% of the target variant can be reliably distinguished from 0%. Tumor biopsy samples used for therapy selection, on the other hand, may contain as little as 5% of the target variant as compared to wildtype sequence. Finally, for noninvasive screening and recurrence applications from peripheral blood, the DNA test must be specific enough to detect mutations at variant allele frequencies of 0.1% or less.

Clinical application of DNA analysis technologies necessarily lags research tool development, because high analytic accuracy is a necessary but not sufficient precondition of a diagnostic test. The FDA requires that IVDs meet the more difficult requirements of high clinical sensitivity and specificity; see Box 3 for an explanation of these metrics. Additionally, market forces generally demand that the test must inform meaningful clinical action, otherwise the test is said to have poor "clinical utility" and will not be reimbursed by payers such as the Center for Medicare and Medicaid Services (CMS). Finally, even an IVD that is FDA cleared/approved and CMS reimbursable face the risk of poor customer adoption. For these reasons, many promising technologies fail to transition to truly impact human health. However, with rising public awareness, government support, and private investment in DNA molecular diagnostics, we envision that an increasing number of clinical DNA detection assays will become available and utilized in the coming years.

In this review, we specifically do not discuss a fourth class of DNA analysis and diagnostic technology, known as isothermal amplification [16]. Common isothermal amplification methods use a polymerase to generate amplicon products templated from an analyte sequence, but rely on enzymes (rather than high temperature) to separate the two strands of the double-stranded amplicon. The advantage of isothermal amplification is that by eschewing precise temperature control equipment, these methods are more suitable in point-of-care and resource-limited settings. However, isothermal amplification generally struggles with precise quantitation, multiplexing, and sequence selectivity; consequently isothermal amplification are rarely used in research and hospital laboratory settings.

2. Polymerase chain reaction

The polymerase chain reaction (PCR) is a method by which a template DNA molecule is amplified using synthetic DNA primers, a DNA polymerase, and dNTPs. The mixture is cycled between at least 2 temperatures: a high temperature for denaturing double-stranded DNA into single-stranded molecules (e.g. 95 °C) and a low temperature for the primer to hybridize to the template and for the polymerase to extend the primer (e.g. 60 °C). Each temperature cycle, in principle, doubles the quantity of target sequence, so even a few copies of a target DNA molecule can be rapidly amplified to nanomolar concentrations,

Box 1
US regulations for clinical use of DNA tests.

In vitro diagnostics (IVDs). The US government defines IVDs as reagents, instruments, and systems intended for use in the diagnosis of disease or other conditions, including a determination of the state of health, in order to cure, mitigate, treat, or prevent disease, and considers them as medical devices [146]. Medical devices are classified by complexity into class I, class II, or class III, with the last being most complex and subject to the most regulatory scrutiny. Most DNA tests are considered class II medical devices (e.g. tuberculosis PCR [147] and cystic fibrosis NGS [148]).
The FDA regulates commercial IVDs for reasonable assurance of safety and effectiveness. To legally commercially sell an IVD requires an FDA pre-market submission, which is one of the following: (1) 510(k) clearance, (2) de novo clearance, or (3) pre-market approval (PMA). FDA 510(k) clearance is the easiest options, but requires that an IVD to show substantial equivalence to a predicate 510(k) cleared device, the latter of which also must have shown substantial equivalence to an earlier predicate device, in a chain that follows back to a product legally marketed before 1976 or to a de novo approved device. De novo clearance may be obtained if no suitable predicate device exists, but the FDA deems the IVD to be low to moderate risk. FDA approval is the highest bar reserved for novel class III IVDs.
Laboratory developed tests (LDTs). Complicating the regulatory process for DNA tests are LDTs, which developed and used at a single laboratory, certified by the Center for Medicare and Medicaid Services (CMS) under the Clinical Laboratory Improvement Amendments (CLIA). There are roughly 250,000 CLIA laboratories in the US [149], and the vast majority of the LDTs offered are low complexity (e.g. blood cholesterol testing) and not reviewed by the FDA.
IVDs are instruments or reagent kits manufactured in one location and subsequently shipped to hospital sites where it produces diagnostic results that directly inform clinical decision. Using a LDT, in contrast, generally consists of a physician or hospital mailing a patient sample to a CLIA lab facility, and receiving after a few days a test report that advises the physician. LDTs, unlike IVDs, do not require clinical validation, though many marketed LDTs have been significantly clinically validated, such as the standard of care OncoType Dx assay by Genomic Health for breast cancer recurrence likelihood prediction [150].
Regulatory uncertainty on the future of LDTs. In October 2014, the FDA proposed a framework for it to regulate LDTs, partially in response to the expanded number of NGS-based cancer-related LDTs. The FDA report was received with varying degrees of skepticism by pathologists, CLIA laboratories, and the CMS. Since then, the College of American Pathologists (CAP) [14] and Association for Molecular Pathology (AMP) have provided comments and/or alternative proposals [15]. As of this writing, the future of LDT regulation remains unclear, though it is likely that LDTs will continue to be allowed in some manner, albeit with more stringent clinical validation requirements.

Box 2
Clinical roles of DNA diagnostics.

To be valuable to society, a DNA diagnostic test must provide information that can potentially affect a clinical decision. Thus, DNA diagnostic tests may be classified by the types of patients the diagnostic appeals to, and the corresponding decisions that it may affect [151].
1. *Risk assessment*. A significant fraction of population bear germ line (inherited) mutations that predispose the individual to a disease. Analyzing people who do not show overt disease symptoms to inform future likelihood of developing a disease is known as risk assessment. For example, the BRCA1/2 test by Myriad Genetics assesses women for lifetime risk of breast cancer.
2. *Screening*. Some diseases have conventional diagnostics that are invasive or inconvenient and not frequently employed without clear disease indication. Analyzing people who do not show overt disease symptoms to detect early stages of a disease is known as screening. For example, the ColoGuard test by Exact Sciences screens people over 55 years old for colorectal cancer.
3. *Diagnosis*. Patients may present nonspecific disease indications (e.g. pain, lowered blood pressure) which may be associated with multiple diseases. A diagnostic test provides definitive disease assessment. For example, the SeptiFast test by Roche diagnoses patients presenting sepsis symptoms for the 25 most common pathogens causing bloodstream infection.
4. *Prognosis*. For diseases such as cancer with relatively long time scales and multiple possible progression trajectories, the prognosis of disease progression can provide valuable information on treatment options considered. For example, the OncoType Dx test by Genomic Health estimates risk of breast cancer recurrence based on the expression levels of RNA within breast cancer biopsy tissue, in order to inform whether a patient should seek chemotherapy.
5. *Therapy selection*. Multiple treatment options may be available for a particular disease with varying efficacies, side effects, and prices depending on the genetics of the patient. Many cancer therapeutics such as tyrosine kinase inhibitors are specifically effective or ineffective against tumors bearing specific mutations. For example, the Foundation One panel by Foundation Medicine analyzes sequence variations in 315 genes of metastatic cancer patients who have failed first-line treatment.
6. *Monitoring*. Following treatment, a patient may go into disease remission, but will be at elevated risk for recurrence. Monitoring post-operation cancer patient disease status via analysis of DNA in peripheral blood is a promising new direction for improving health outcomes. Although there are not any products widely validated and adopted at the moment, several companies such as Sysmex Inostics and Guardant Health have announced intentions of developing cancer recurrence tests based on low-level mutations.

which can be subsequently detected via fluorescence or other means. PCR is currently the most widely used method for detection of DNA sequences [17].

Compared to the two other classes of technologies reviewed, PCR's main strengths are accurate quantitation, high molecular sensitivity, and ease of use. Quantitative PCR, for example, is used as a gold standard for DNA and RNA quantitation that is generally considered to be more accurate than either microarrays or NGS. PCR's main weakness is its inability to perform highly multiplexed assays, due to primer dimer formation that result in false positives or false negatives.

2.1. ARMS and related technologies

Detection of sequence variations using PCR typically involves the design and use oligonucleotide reagents (i.e. primers and blockers) that amplify the variant of interest more efficiently than the corresponding wildtype DNA sequence. The amplification-refractory mutation system (ARMS) is an early method for detecting DNA sequence variants, including single nucleotide variants [18]. The operating principle behind ARMS is that the enzymatic extension activity of DNA polymerases is highly sensitive to mismatches at or near the 3′ end of the primer-

Box 3
Assay performance metrics.

The terms specificity and sensitivity are used, somewhat confusingly, to describe several different measures of DNA assay performance. In the earliest proof-of-concept stages in the research laboratory, the molecular sensitivity of a DNA sequence variant detection assay typically refers to the concentration or number of molecules of the target DNA sequence that can be unambiguously detected (e.g. 1 fM; 20 copies). The molecular specificity refers to the degree in which the desired DNA sequence variant produces a signal higher than the wildtype or other variants (e.g. quotient of observed signals for positive and negative control samples). In NGS, molecular specificity is closely related to the intrinsic error rate of sequencing.

The evaluation of a test's analytical sensitivity and analytical specificity assumes a set of positive and negative control samples bearing and lacking the DNA sequence variant(s) of interest, respectively. These control samples are often provided by a third party. Analytical sensitivity is the percentage of positive control samples that are correctly assayed as positive, and analytical specificity is the percentage of negative control samples that are correctly assayed as negative. In general, the analytical sensitivity and analytical specificity of a test must be optimized to be very close to 100% before the test is considered for translation into a diagnostic test.

Clinical sensitivity and clinical specificity consider the effectiveness of the test in detecting the disease in patients: clinical sensitivity is the percentage of disease-positive patients that are correctly tested as positive, and clinical specificity is the percentage of disease-negative patients that are correctly tested as negative. Because clinical sensitivity and specificity do not account for low-level details such as DNA target sequences, these metrics can be broadly compared among tests using very different approaches. In general, the FDA requires clinical sensitivity and specificity data for any submission.

Another way to look at the clinical versus analytical metrics is through consideration of content and platform. Content refers to the target genes or variations as that are being detected as indicators of a particular disease status, and platform refers to the instrument and reagents that perform the detection process. Analytic sensitivity and specificity show the performance of the platform, while clinical sensitivity and specificity show the overall performance of both the content and the platform. Depending on the application, either content or platform may be the bottleneck for diagnostic test performance.

template duplex (Fig. 2a). The ARMS primer is designed such that the 3′-most base hybridizes to the target sequence variant, but not the corresponding wildtype sequence.

For single nucleotide polymorphism (SNP) genotyping application where the variant allele is present at 50% or 0% frequency, ARMS primers typically provide sufficient molecular specificity for reliable detection [19,20]. However, for somatic mutation detection from biopsy samples in which the variant may be present at as low as 5% allele frequency, ARMS primers do not consistently provide sufficient sequence discrimination, because the molecular specificity of ARMS primers varies for different target and wildtype sequences. Some mismatches are either more thermodynamically destabilizing or more easily recognized by the DNA polymerase enzyme, and result in low false positive amplification of wildtype sequences. This problem is somewhat mitigated by the fact that DNA is double-stranded, and either strand may be used as the detection target. Alternatively, mismatches may be introduced in the ARMS primer near the 3′ nucleotide to improve specificity, at the cost of reduced PCR yield (Fig. 2b).

Many different companies have developed diagnostics tests based on ARMS PCR primers. Qiagen therascreen [21] and Roche cobas [22] have developed FDA approved PCR tests for detecting lung and colon cancer mutations in the KRAS and EGFR genes; these IVD kits were validated on genomic DNA extracted from formalin-fixed, paraffin-embedded (FFPE) tissue. Biomerieux THxID [23] has developed FDA approved PCR tests for detecting mutations in the BRAF gene for metastatic melanoma, likewise validated on FFPE tissue. AmoyDx is a Chinese company that has developed ARMS primers for a large number of cancer genes, and obtained CFDA approval as well as CE-IVD marking.

A significant improvement to the ARMS primer technology is the use of blocking oligonucleotides in the allele-specific blocker PCR (asbPCR) [24,25] and competitive allele-specific Taqman PCR (castPCR) [26] assays (Fig. 2c). The blocker is an oligonucleotide that hybridizes perfectly to the wildtype template, and thereby suppressing unintended hybridization of the ARMS primer to the wildtype. The blocker is typically functionalized with a chemical moiety (e.g. a minor groove binder, MGB) at the 3′ end to prevent polymerase extension, and optionally to improve binding stability. Applied Biosystems (now part of Thermo Fisher) developed and markets castPCR kits for 586 mutations in 45 cancer genes; the assays are research use only (RUO) and have not been reviewed by the FDA [27].

An alternative approach to allele-specific PCR is to use two-segment primers, such as the dual-priming oligonucleotide (DPO) by Seegene [28,29] and the myT primers by Swift Biosciences [30](Fig. 2d). Both of these primers include a longer 5′ region that primarily contributes hybridization stability and a shorter 3′ region that primarily contributes specificity. The DPO and myT primers are more specific than ARMS primers because a single nucleotide mismatch has a larger thermodynamic effect on the binding stability of a short oligonucleotide region than on a longer region. Seegene recently obtained FDA approval on its Herpes Simplex Virus test and also developed several CE-marked assays for infectious disease diagnostics, such as the Seeplex Diarrhea ACE detection kit [31].

2.2. Blocker PCR

An alternative set of approaches to PCR detection of sequence variant relies on suppression of wildtype amplification through the use of blocker oligonucleotides. In these schemes, the primers are typically not allele specific, and in the absence of the blocker hybridized to the wildtype, the primer would amplify both the variants and the wildtype with roughly equal efficiency. Blocker PCR exhibit two primary benefits over ARMS: first, it is hypothesis-free over the blocker binding region; the sequence of the variants do not need to be known a priori. Second, it offers compounded specificity through multiple cycles of PCR, because the primer does not itself incorporate the polymorphic nucleotide(s). In contrast, ARMS primers are specific only until the first spurious extension event generates an amplicon bearing the sequence variant allele, and thus are more prone to stochastic errors

The first reported use of blocker PCR uses a peptide nucleic acid (PNA) blocker (Fig. 3a) [32]. The DNA polymerase is unable to displace or digest the PNA blocker, so primer extension of the wildtype halts where the blocker binds. The anneal cycle temperature is set such that the blocker does not bind favorably to sequence variants, due to the mismatch bubble formed. The same logic can be applied to different blocker molecule types: DiaCarta's xeno nucleic acids (XNA) are modified nucleic acids with greater binding affinity than PNA and exhibit improved variant discrimination [34], and Biocept's Selector assay using 5′ phosphorothioate-modified DNA blockers as an economical alternative solution [35] (PNA and XNA are expensive to synthesize). The PNAClamp assay by PNA Bio has been CE-marked for IVD use in Europe for guiding cancer treatment based on mutation analysis of the EGFR, KRAS, BRAF, PI3K, and IDH1 genes [33].

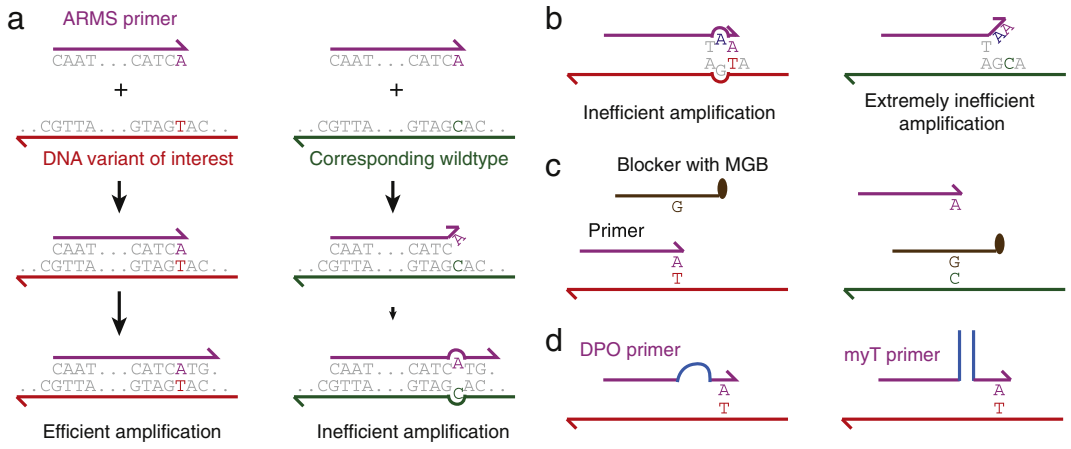

Fig. 2. Detection of sequence variants using allele-specific PCR primers. (a) ARMS primers designed to detect a particular sequence variant. The 3′ nucleotide hybridizes perfectly to the target variant template, but is mismatched to the wildtype template. Taq or other DNA polymerases used in PCR extend the primer off the wildtype with significantly lower efficiency as compared to the target. (b) Intentional introduction of mismatched nucleotide at the penultimate 3′ position. SNV amplification efficiency is reduced, but wildtype amplification is almost completely inhibited. (c) ARMS primer with wildtype blocker, as used in asbPCR and castPCR. The blocker competes with the primer in hybridizing to the wildtype template, and thus further suppresses unintended amplification of the wildtype. Due to the relative hybridization thermodynamics, the blocker does not significantly impact primer binding or extension for the target template. (d) DPO and myT primers are two-segment primers with a longer 5′ binding region and a shorter 3′ binding region. The segments are connected by a poly-inosine linker for DPO primers, and by an orthogonal double-stranded DNA region for myT primers.

A variation of the blocker PCR approach is the co-amplification at lowered denaturation temperature PCR (COLD-PCR) [36,37] and the related ICE COLD-PCR [38,39] assays. These assays rely on more complex temperature cycling protocols to kinetically favor the hybridization of the blockers to the wildtype templates. The MX-ICP assays by Transgenomic is based on ICE COLD-PCR and is offered in the United States as an LDT.

Integrated DNA Technologies developed a conceptually different type of blocker PCR, known as RNAse H-dependent PCR (rhPCR) [40]. Unlike the other blocker PCR implementations, the blocker is not a distinct molecule, but rather a 3′ region of the primer including an RNA nucleotide at the polymorphic site and a 3′moiety that inhibits polymerase extension (Fig 3b). When the primer/blocker binds to the desired sequence variant, the RNA nucleotide is paired to its complement on the template, and is cleaved by a temperature robust RNAse H2 enzyme. The cleaved primer is subsequently extended by the DNA polymerase. When the primer/blocker binds to a wildtype sequence, the RNA nucleotide is mismatched, and is not cut by the RNAse H2 enzyme. One advantage of the rhPCR technology over other blocker PCR assays is that it suppresses primer dimer formation and nonspecific genomic amplification, due to the enzymatic action of the RNAse H2 enzyme.

2.3. Multiplex PCR

All PCR technologies described above struggle to variant extents with multiplexing, the simultaneous analysis of multiple target sequence variants. There are three main difficulties in multiplexed PCR: the depletion of dNTPs by the highest concentration amplicons, the orthogonal readout of different amplicons, and the formation of primer dimers during amplification

In a homogeneous PCR reaction, the dNTPs used for primer extension become depleted as the amplicon accumulates. In multiplex PCR amplification of several targets, the presence of one high concentration target can effectively suppress the amplification of other targets.

Real-time PCR (a.k.a. quantitative PCR, qPCR) requires the use of a fluorophore to indicate amplicon concentration at different cycles; the cycle at which the amplicon concentration exceeds a threshold is known as the quantitation cycle (Cq), which is log-linearly related to the initial target concentration. The number of spectrally distinct fluorophores limits the number of targets that can be simultaneously quantitated [41]. The traditional limit is roughly 5 fluorophores [42,43].

Primer dimers refer to the unintended interaction between primers that result in the formation of short amplicons with sequence unrelated to any templates. For single-plex PCR with 2 primers, careful design and optimization of primer sequence can result in a good set of primers with little primer dimer formation. However, in multiplex PCR for simultaneous analysis of N templates, there are 2N primers, which result in at least $4N^2$ possible primer dimer interactions. In reality, because there are complex primer dimer mechanisms involving three or more species, the complexity scales even worse with the number of primers.

The engineering solution to the multiplex PCR problem is the development of instruments and disposable chips that compartmentalize the PCR reaction, so that there is only one set of primers in each compartment. Biofire Diagnostics (now part of Biomerieux) developed the FilmArray system to allow simultaneous PCR analysis of 10's of template sequences [44]. Biofire has multiple FDA cleared diagnostic panels including for respiratory and gastrointestinal infectious diseases. Cepheid, another developer of multiple FDA cleared/approved infectious disease diagnostics instruments and kits, announced in 2012 the development of an instrument capable of 1000-plex PCR analysis.

2.4. Digital PCR

Reliable detection and quantitation of low allele frequency variants by conventional PCR remains challenging. Despite the best molecular primer and blocker designs, there will inevitably still be some degree of false amplification of wildtype sequence, due to the stochastic nature of molecular interactions. For sequence variant detection in particular, allele frequency quantitation is complicated by the different per-step amplification yields

In digital PCR, a single-plex PCR reaction is split into thousands to millions of droplets [45,46,47] (Fig. 4) Typically, the droplets are formed by mixing the aqueous reaction with oil to form an emulsion (Biorad ddPCR, Rain Dance RainDrop), although microfluidic approaches are also available (Thermo Fisher Open Array, Fluidigm BioMark). The benefit of digital PCR is that with so many droplets, in each droplet, there is expected to be 0 or 1 copies of the target sequence. Because 0 and 1 copies of the target sequence result in a large difference in signal, positive amplification droplets can be easily distinguished from droplets lacking template. Quantitation is also facilitated because the number of positive amplification droplets can be directly counted.

Currently, digital PCR is used for academic and clinical research purposes, and also used clinically as laboratory developed tests. For example,

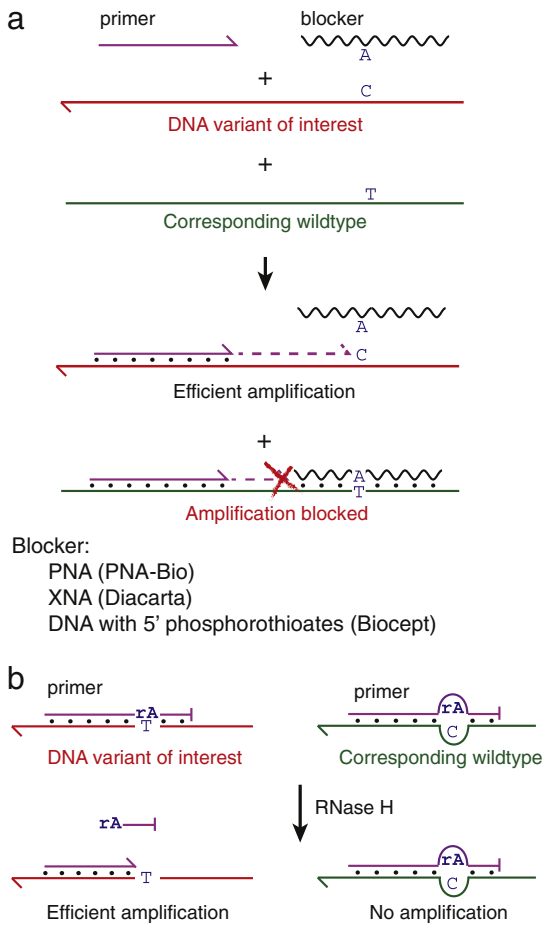

Fig. 3. Allele-specific variant amplification through blocker PCR. (a) In blocker PCR, the primer binds upstream of the polymorphic locus to be queried. The blocker binds favorably to the wildtype template, but not to any variants. Polymerase extension of the primer halts when it reaches the blocker on the wildtype template. The identity of the blocker oligonucleotide may be PNA (PNAClamp), XNA (DiaCarta), or DNA with 5′ phosphorthioate modifications (Biocept). (b) In rhPCR, the primer is functionalized internally with an RNA nucleotide that corresponds to the target allele, and at the 3′ with an nonextensible moiety (typically a 3-carbon spacer). When the primer hybridizes to the target sequence variant template, a temperature robust RNAse H2 enzyme cleaves the primer to the 5′ of the RNA base, and the deprotected primer can subsequently be extended. The RNAse H2 enzyme does not efficiently cleave the RNA nucleotide when it is mismatched (in the primer-wildtype complex).

the Trovagene PCM V600E assays analyze BRAF mutations from cell-free DNA in urine samples using digital PCR.

3. Hybridization

Hybridization is the process by which a synthetic DNA probe or primer binds (via Watson–Crick base pairing) to a biological DNA target sequence. Hybridization forms the basis of all modern DNA analysis and diagnostic techniques, but in the absence of either DNA amplification or signal amplification, hybridization does not provide sufficient molecular sensitivity for practical use. More commonly, hybridization is used in conjunction with PCR or with fluorescence microscopy. Recent advances in sensor technologies may allow hybridization in the absence of enzymatic DNA amplification to be a viable alternative to PCR and NGS.

Compared to the two other classes of technologies reviewed, hybridization's main strengths are its simplicity, multiplexing, and robustness. Because hybridization is a biophysical phenomenon, it proceeds in many buffer conditions, unlike enzyme-based assays with narrow acceptable buffer compositions. Hybridization's main

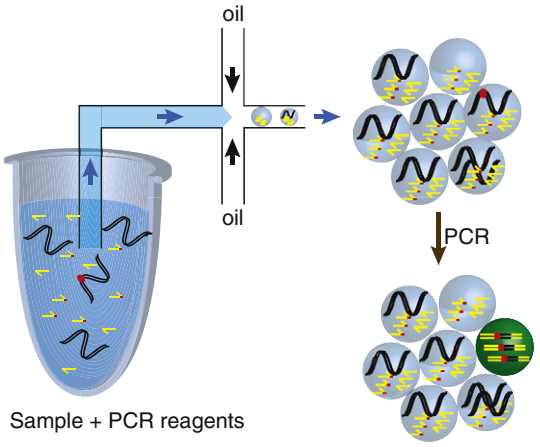

Fig. 4. Digital PCR. An aqueous solution containing the DNA template sample and PCR reagents are microfluidically mixed with oil to generate nanoliter to picoliter sized droplets. Each droplet acts as an individual PCR reaction, and there is one or only a few template molecules in each droplet. Thus, differences in amplification trajectories for the target variant and wildtype template are more pronounced.

weakness is that it does not provide sequence amplification, and must be paired with either signal amplification technology, or a highly sensitive readout instrument.

3.1. Microarrays

Microarrays use spatial arrangement to solve the multiplex readout problem (Fig. 5). DNA probes of different sequences are functionalized onto a surface at different positions. A nucleic acid sample containing targets of interest are 3′ fluorophore-labeled using a terminal transferase, and then hybridized to the microarray [48,49]. The positions of the fluorescent spots indicate the identities of the targets detected, and the fluorescence intensity indicates quantity

Microarrays may be used for the direct (unamplified) detection of high expression RNA species from large sample volumes [50,51], and can also be applied to amplicons from a multiplex PCR reaction [52,53]. The molecular sensitivity of microarrays is limited by the hybridization efficiency of labeled targets to the microarray, as well as the sensitivity of the fluorescence microscope used for imaging and the autofluorescence of the microarray chip. More than a million of different probes can be synthesized on an array with Affymetrix's Genechip technology with a detection limit of one to ten copies of mRNA per well.

In principle, microarrays should provide highly quantitative information regarding nucleic acid concentration. In practice, however, there is substantial quantitation bias across different genes and transcripts, across different microarray platforms, and even across different microarray chips by the same manufacturer [54]. First, hybridization yield and kinetics are nonlinearly affected by the density of probes on the surface: probe molecules hybridize nonspecifically to other probe molecules at high density, to various extents based on sequence. Second, the lengths and sequence of the target molecules affect hybridization kinetics. Third, the quantum yield of fluorophores are known to be affected by both neighboring DNA sequence and by proximity of other fluorophores. Consequently, optimized microarrays are typically considered to produce repeatable relative quantitation of different nucleic acid targets [55], rather than absolute concentration.

A number of FDA approved or cleared diagnostics used microarrays: Agendia's MammaPrint assays [56] for breast cancer recurrence profiles the expression of 70 genes to inform breast cancer recurrence risk, Autogenomics INFNITI CYP2C19 assay [57] profiles genetic polymorphisms that impact therapeutic response to antidepressants and antiepileptics, and the Affymetrix's CytoScan Dx [58] evaluates developmental

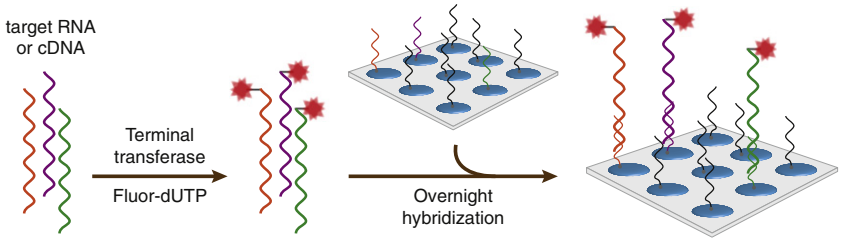

Fig. 5. DNA microarrays for profiling RNA expression. The RNA or cDNA to be profiled is fluorophore-labeled using a terminal transferase enzyme. The labeled targets are then hybridized to the microarray, a surface functionalized with different probe sequences at different positions. Fluorescent intensity at a particular position indicates the concentration of its corresponding labeled target.

delay, intellectual disabilities, and congenital anomalies based on chromosomal mutation analysis. Additionally, Agilent is seeking FDA clearance for its SurePrint gene expression microarrays.

3.2. Fluorescent barcodes

Fluorescence barcodes collectively comprise an alternative approach to highly (100 to 1000) multiplexed readout. Fluorescent barcodes can generally be divided into two flavors: intensity barcodes, or geometric barcodes (Fig. 6). Intensity barcodes use the absolute or relative intensities of several fluorophores to indicate sequence identity. Fig. 6a shows the Luminex xTag approach to intensity barcoding: each silica particle is functionalized with different number of molecules of two spectrally distinct fluorophores, with the intensities of the two fluorophores indicating the species identity; a third fluorophore functionalized to an amplicon indicates the presence of the species [59]. Assuming that each fluorophore intensity can be distinguished to 30 levels, the Luminex approach allows up to 900 barcodes. Luminex developed an FDA approved respiratory disease diagnostics panel based on its fluorescent barcode technology

Geometric barcodes use the orientation of spectrally distinct fluorophores to convey sequence identity; Fig. 6b shows the Nanostring approach of electrophoretically stretching out a nucleic acid barcode. The barcode allows 6 distinct spots, each with 4 possible fluorophore colors; even restricting that neighboring spots must be distinct fluorophores (to ease imaging processing), there are 972 possible barcodes [60,62]. Prosigna is Nanostring's FDA cleared panel for predicting breast cancer recurrence based on the measured expression levels of 50 genes.

Another approach to geometric barcodes taken by Firefly Bioworks (now subsidiary of Abcam) is a physical patterning on a micron-scale hydrogel particles [61]. Because of the large size of the barcodes and advances in micro- and nanofabrication, the number of different potential barcodes is orders of magnitude greater than prior approaches. Abcam is applying these hydrogel barcodes to microRNA profiling research applications.

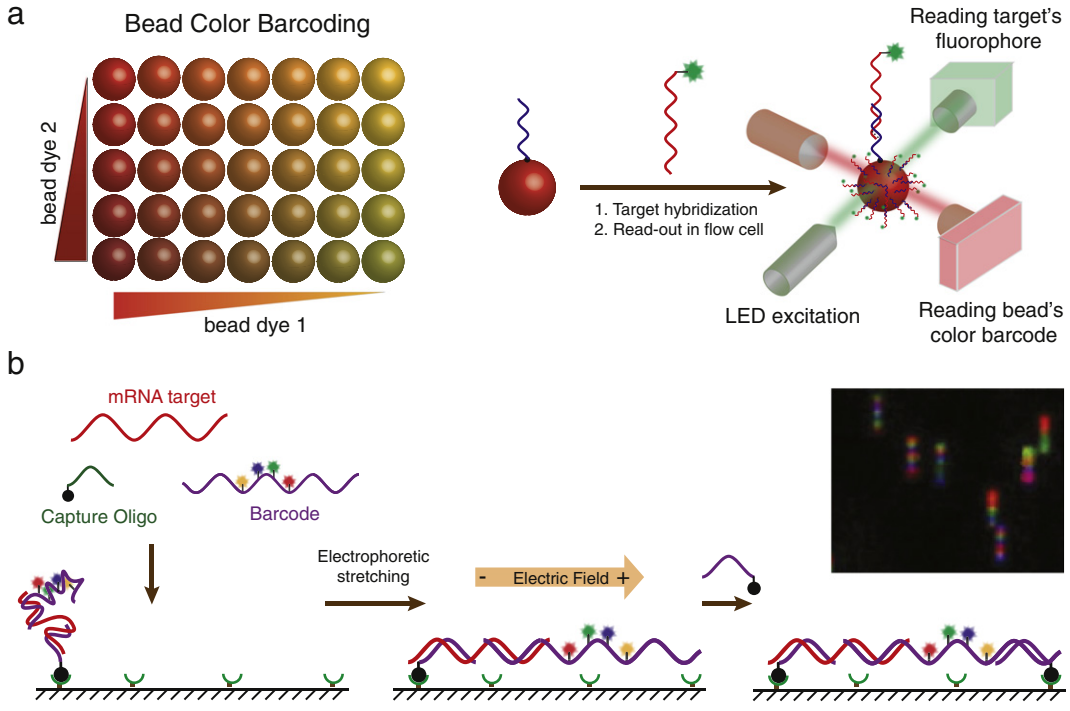

Fig. 6. Fluorescent barcodes for hybridization-based nucleic acid analysis. (a) Luminex xTag intensity barcodes. Beads are functionalized with two distinct fluorophore molecules, as well as a probe oligonucleotide. The intensities of the two bead fluorophores indicate the sequence identity of the functionalized probe oligonucleotide. Hybridization of a labeled target or amplicon sequence is detected via a third fluorescence channel. (b) Nanostring nCounter geometric barcodes. Target RNA molecules colocalize a fluorescent barcode probe with a biotinylated capture probe, and subsequently deposited on a surface. The target and fluorescent barcode are electrophoretically stretched based on the negative charged of the DNA backbone, resulting in a visible linear chain of fluorescent spots.

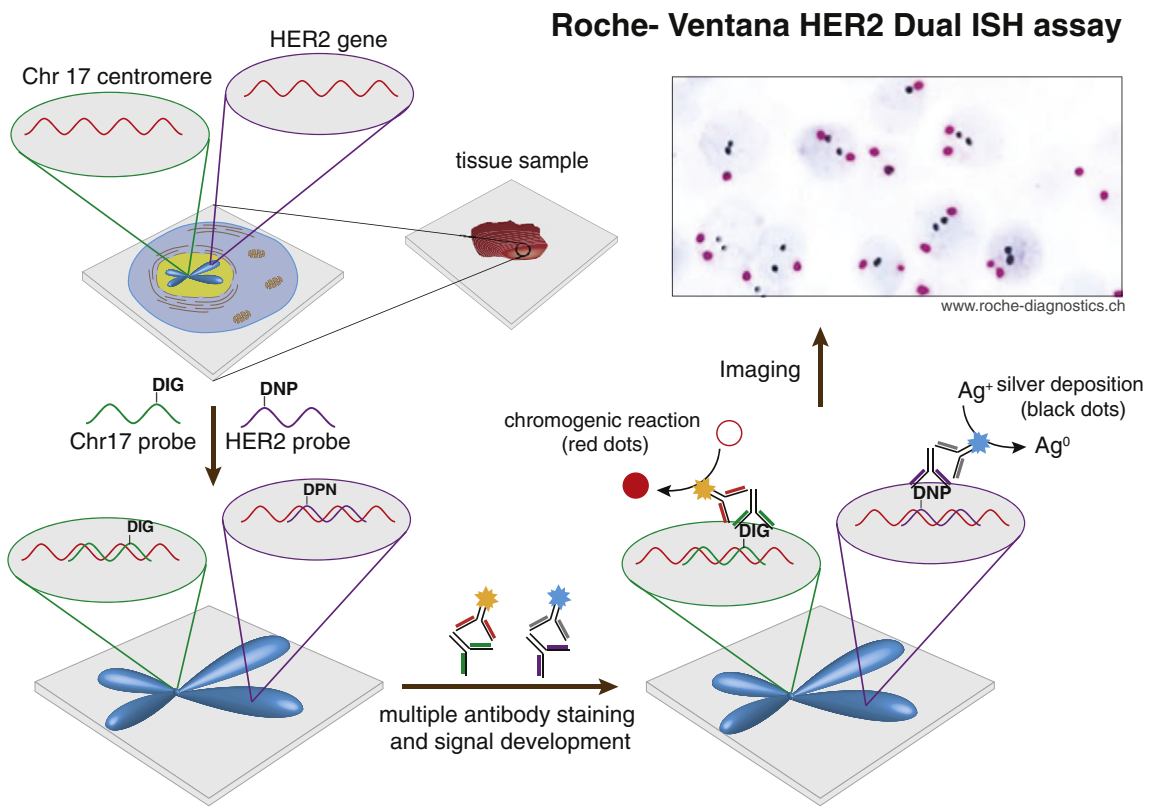

Fig. 7. In-situ hybridization. Shown here is the Roche Ventana HER2 amplification assay for determining herceptin efficacy. DNA within fixed tissue sections are hybridized to hapten-labeled probes, which subsequently recruit antibodies that effect chromogenic signal observable by light microscopy. DIG (digoxygenin) and DPN (dinitrophenyl) are two orthogonal haptens used to differentially label the chromosome 17 centromere and the HER2 gene, respectively.

3.3. In situ hybridization

In situ hybridization (ISH) seeks to provide not only sequence and concentration information regarding target genes and variants of interest, but also spatial positioning of the targets within its native tissue [63–65]. This renders ISH particularly suitable for analysis of copy number variations (CNVs) in heterogeneous cell or tissue samples, because the relative increase in signal within affected cells is far greater than averaged over the entire sample

In ISH, the nucleic acids to be imaged are first fixed to the protein matrix of the cells, typically using formaldehyde (a.k.a. formalin) or methanol, to prevent diffusion (Fig. 7). Subsequently, DNA or RNA oligonucleotide probes are introduced and allowed to hybridize to the fixed target nucleic acids; unbound probes are washed away. Finally, an imaging agent is introduced, the exact identity of which depends on the assay. For fluorescent ISH (FISH), the DNA probes are themselves labeled with fluorophores, or are modified with a hapten (e.g. digoxygenin, DIG) that recruits a fluorescent antibody. For chromogenic ISH (CISH) or silver-enhanced ISH (SISH) [68], the hapten-labeled DNA probe recruits an antibody, which recruits a secondary antibody functionalized with an enzyme that generates the desired signal. Compared to CISH [66,67] and SISH, FISH possesses lower molecular sensitivity, but allows simultaneous multiplex imaging to 3–5 different species [69,70]. Additionally, directly labeled FISH probes could bleached or washed away to allow the same sample to be repeated imaged for different genes, further improving multiplexing.

ISH may be applied to both DNA and RNA targets. Thus far, clinical usage of FISH has focused on DNA copy number variation: Roche Ventana and Abbott both developed FDA approved ISH assays for detecting HER2 amplification to inform herceptin efficacy for breast cancer patients. However, Advanced Cell Diagnostics has recently announced intention to pursue HPV diagnostics via its RNAscope FISH assay [71].

3.4. Other readout modalities

Alternative technologies for readout of nucleic acid targets or amplicons have been developed that show high molecular sensitivity. These include electrochemical readout (e.g. Xagenic [72,73]), magnetic resonance (e.g. T2 biosystems [74,75]), nanoparticle aggregation-induced optical scattering (e.g. nanosphere [76]), and chemiluminescence (e.g. Hologic HPA [77]); several of these have obtained FDA clearance or approval as infectious disease IVDs. With these sensors, the need for many cycles of PCR amplification to increase amplicon concentration is mitigated or eliminated. However, these devices generally show limited molecular specificity, and thus are more appropriate for detecting the presence or absence of pathogen genes, rather than sequence variations.

4. Next-generation sequencing

Next generation sequencing (NGS) is a family of approaches for massively multiplexed sequence analysis of DNA and RNA. Unlike traditional Sanger sequencing, which requires a homogeneous DNA template as input, NGS allows analysis of heterogeneous samples, and simultaneously provides sequence information for more than 10 million randomly selected nucleic acid molecules in a sample [78,79]. Because of the large number of reads, NGS is uniquely suitable for nucleic acid analysis and diagnostic requiring multiplexed analysis of many genes and their variants.

No chemistry is perfect, and all NGS platforms suffer a finite intrinsic error rate due to signal ambiguity, enzyme infidelity, imperfect deprotection, etc. Sequencing errors complicate the calling of variants, especially low frequency ones. Recent innovations in molecular barcoding [80,81] have significantly reduced the NGS error rates, at the cost of increasing sequencing depth. Although there are not currently

any FDA cleared or approved NGS assays for cancer-related diagnostics, the Foundation One LDT has garnered interest and usage from the clinical oncologists.

4.1. Mainstream NGS platforms

4.1.1. Illumina

The Illumina NGS system [82] is based on the idea of sequential fluorophore-labeled nucleotide base addition combined with fluorescence imaging (Fig. 8a). This platform relies on "bridge PCR" on a surface using tethered primers to generate local "polonies" of amplicons. All amplicons within each polony should have the same sequence, and generate the same colored fluorescence during each nucleotide incorporation cycle (Fig. 8a). The color of the fluorescence corresponds to the identity of the incorporated nucleotide. Bridge PCR [83,84] also uniquely allows a DNA fragment to be sequenced from both ends (pair-end sequencing). Pair-end sequencing improves final read alignment, insertion and deletion calling, rearrangement identification, and FASTQ quality score [85–87].

Illumina is currently the leader both in sequencing error rate and in sequencing cost per read. The sizes of the polonies (200 nm diameter) define the fundamental limit of throughput for a single flow cell, rendering Illumina the highest throughput NGS platform today (1.5 Tb). Illumina has continued to invest in flow cell technologies to increase data output and quality, with a recently released patterned flow cell in which immobilized primers are arranged in a defined surface array allowing for a precise control of consecutive cluster generation [88,89]. Currently, 10^6 Illumina sequencing reads costs roughly $2.

On the other hand, the Illumina platform is also the slowest of the NGS platforms reviewed here, with each cycle (nucleotide incorporation and imaging) taking roughly 5 min. This is because both fluorophore cleavage and the high resolution flow cell imaging steps are time consuming. For a 300 nt paired end NGS run, the sequencing takes 2 full days. Improvements in nucleotide chemistry and/or microscopy are unlikely to significantly reduce the sequencing run time.

For diagnostic applications, Illumina has developed benchtop sequencing platforms (MiSeq and NextSeq) with the lower throughput (<15 gigabases). The MiSeqDx system is the first NGS platform to receive FDA clearance for IVD use [90]. Thus far, two FDA cleared NGS assays have been developed to function on the MiSeqDx platform, both for cystic fibrosis variant genotyping [91,92]. In addition, many LDTs (e.g. Foundation Medicine's FoundationOne panel) use Illumina sequencing to inform cancer therapy.

4.1.2. Ion Torrent

The Ion Torrent sequencing platform uses pH rather than fluorescence for nucleotide identity readout [93]. During primer extension, a proton ion (H^+) is released for each nucleotide incorporation event (Fig. 8b). The released protons cause a localized and transient pH change that is detected by a miniature pH sensor (ion-sensitive field-effect transistor, ISFET) on a matrix of CMOS (complementary metal-oxide-semiconductor) elements. In each sequencing step, only a single

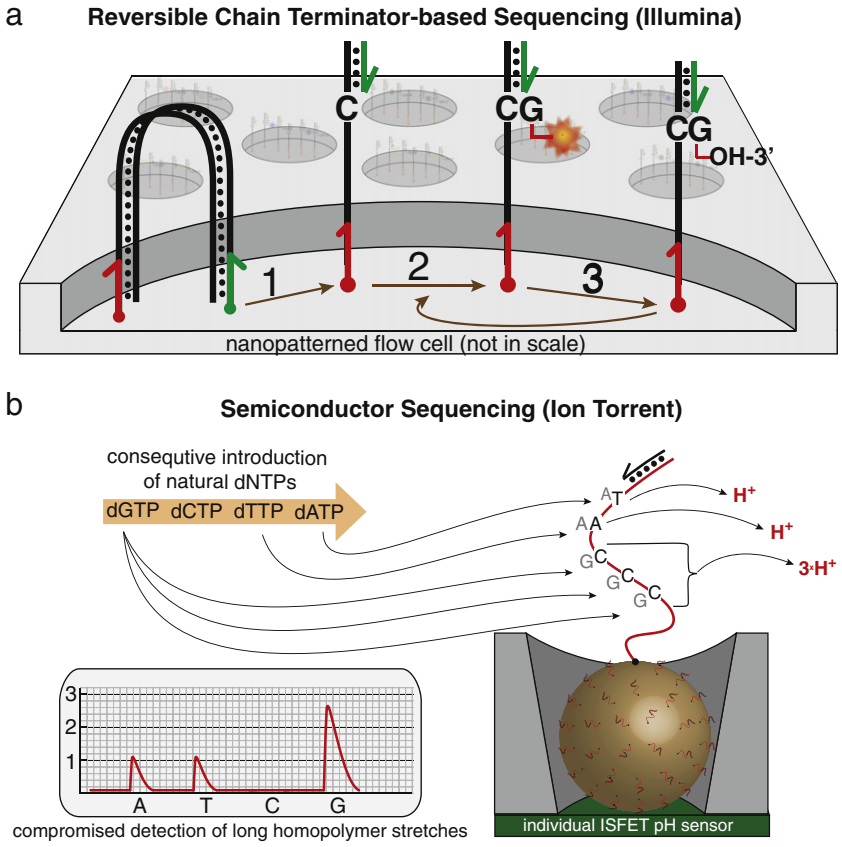

Fig. 8. Mainstream NGS platforms. (a) Illumina sequencing using bridge PCR and repeated fluorescent imaging of polonies. The target DNA is appended with adaptors at both 5′ and 3′ ends, and these adaptors allow the target to be captured and amplified on the surface of a patterned flow-cell. During the actual sequencing, in each cycle all dNTPs, each with a distinct fluorophore label, is added simultaneously and the each polony incorporates one nucleotide to produce a bright fluorescent spot detected by imaging. Chemical treatment cleaves the fluorophore and blocker moiety, allowing a new cycle of nucleotide incorporation to begin. (b) Ion Torrent sequencing using pH to sense nucleotide incorporation. During polymerase extension and nucleotide incorporation, a proton is released that transiently changes the pH of a microwell. By flowing in one type of unmodified dNTP (e.g. dATP) at a time, the miniaturized ISFET pH sensors report the microwells incorporated the introduced nucleotide. Homopolymeric sequences (e.g. AAAAA) lead to release of a larger number of protons that may be difficult to precisely quantitate.

nucleotide type (e.g. dATP) is introduced into the reaction flow cell; only the DNA fragments with the corresponding nucleotide as the next base will show a pH signal. All four nucleotides are cycled in this fashion to perform the sequencing.

The nucleotides incorporated in the Ion Torrent NGS workflow are unmodified nucleotide triphosphates, so there are no chemical deprotection steps. Additionally, the pH sensors react nearly instantaneously. Consequently, Ion Torrent sequencing is significantly faster than Illumina sequencing, taking roughly 3 h for 300 nt reads. Another advantage of Ion Torrent NGS platforms is that the instruments themselves are significantly less expensive than corresponding Illumina instruments, due to the relatively lower cost of pH sensors compared to optical readout systems. Finally, Ion Torrent's NGS platform and its upstream Ampliseq protocol allow analysis of biological DNA samples of down to 10 ng (compared to Illumina's 100 ng).

The major disadvantages of Ion Torrent sequencing are its relatively higher intrinsic error rate and its higher per read cost as compared to Illumina. The higher error rate arises both from homopolymer regions that generate pH signal that is difficult to accurately quantify past 4 nucleotides (e.g. AAAAA), and from increased enzymatic misincorporation rates when only one nucleotide triphosphate is present. Currently, 10^6 Ion Torrent sequencing reads costs roughly $10.

Ion Torrent's PGM-Dx instrument has received FDA clearance for clinical use. In contrast to Illumina, Ion Torrent has not released any of disease specific assays, but decided to promote its PGM-Dx system as an open platform allowing clinicians to develop and validate clinical assays [94].

4.2. Alternative NGS platforms

4.2.1. Pacific Biosciences

Whereas both Illumina and Ion Torrent NGS platforms have read lengths limited to roughly 300 nt, Pacific Bioscience's SMRT NGS platform allows average read lengths of 10,000 nt and maximum read lengths of roughly 50,000 nt. This feature renders PacBio uniquely suitable for high quality de novo genome assembly [95,96], isoforms profiling [97], and structural variants resolution [98,99].

SMRT sequencing is based on real-time observation of nucleotide incorporation on a growing DNA chain (Fig 9a) [100]. Incorporated nucleotides are fluorescently labeled in the gamma-phosphate position, so that they are naturally cleaved during the incorporation process. Thus, unlike Illumina sequencing, SMRT sequencing does not pause after each nucleotide incorporation for chemical cleavage and fluorescence imaging. Another technology, known as the zero-mode waveguide, allows SMRT sequencing to continuously sense fluorophores only near the surface-bound polymerase, reducing background signal of unincorporated nucleotides.

Despite the ingenuity of the employed technologies, SMRT sequencing currently exhibits intrinsic error rates far worse than Illumina or Ion Torrent. One solution that PacBio developed to mitigate the sequencing error problem is to circularize DNA targets to allow repeated sequencing. By sequencing each nucleotide multiple times on both the sense and antisense strands, error rates can be statistically improved. However, obtaining sufficient single-molecule read depth to eliminate sequencing error would limit the length of the DNA target, and wipe out PacBio's primary competitive advantage. SMRT sequencing also is significantly more expensive (per read) than both Illumina and Ion Torrent. Currently, 10^6 PacBio sequencing reads costs roughly $300. In 2013, PacBio and Roche announced a partnership to pursue clinical IVD development.

4.2.2. Oxford Nanopore

Oxford Nanopores NGS approach differs from other technologies described here in that it does not rely on polymerase extension of DNA primers. Instead, the DNA target molecules are threaded and pulled through an enzyme nanopore embedded in a synthetic polymer membrane (Fig. 9b). The instantaneous electric current through the membrane is affected by the identity of the nucleotides in the pore complex at that moment, and can thus be used to inform sequence [101,102].

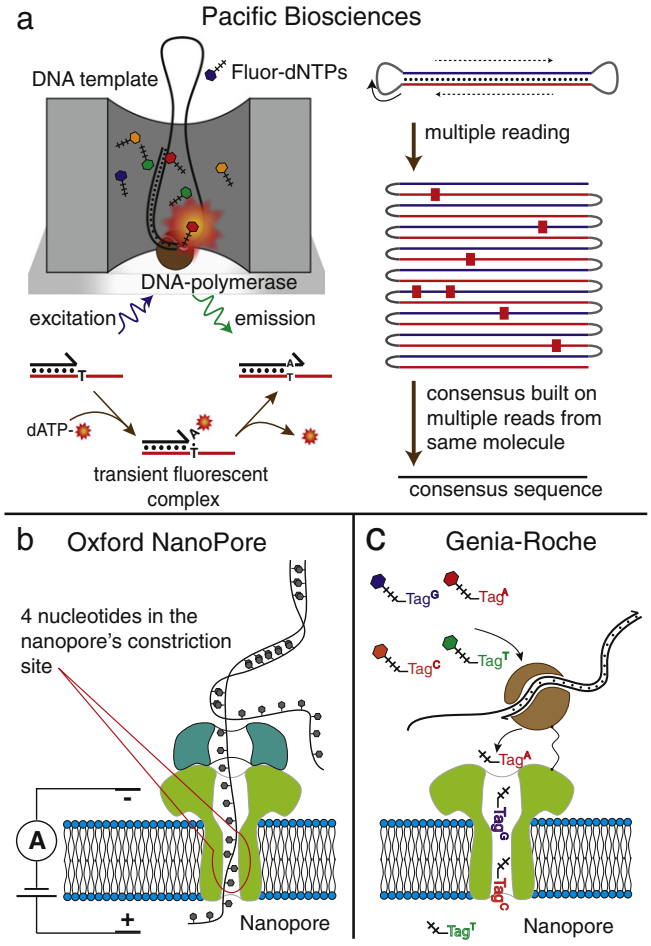

Fig. 9. Single molecule sequencing approaches. (a) Pacific Biosciences Single Molecule Real Time Sequencing (SMRT). A polymerase is tethered to a surface near a zero mode waveguide (ZMW) that sensitively detects fluorescence in a very narrow zone near the surface. As fluorophore-labeled dNTPs are incorporated into a growing amplicon, the fluorophores themselves are cleaved off and produce a transient burst of fluorescence detected by the ZMW. The DNA templated may be circularized to allow repeated sequencing of the same or complementary regions, reducing sequencing error. (b) Oxford Nanopore sequencing. Target DNA molecule is dragged through the nanopore using electric current via a special motor enzyme. The 4 nucleotides in the nanopore affect the instantaneous electrical current across the nanopore; this current is used to infer sequence using a Hidden Markov model. (c) Genia uses a DNA polymerase tethered to one side of the nanopore, and PEG-modified dNTPs. As the polymerase copies the target DNA sequence, the PEG tags are released and forced through the nanopore. The size of the PEG tag corresponds to the identity of the incorporated nucleotide, and determines the instantaneous current across the nanopore when it is in transit through the nanopore.

The size of the pore complex determines the number of nucleotides that simultaneously fit inside and affect the electrical current [103,104]; it is easier to determine sequence information from smaller pore complexes with fewer transit nucleotides. Oxford Nanopores current pore proteins are derived from *Mycobacterium smegmatis*, and fits 4 nucleotides [105,106]; this means that a $4^4 = 256$ state Hidden Markov model is needed to deconvolute current to resolve sequence [107]. The large number of states contributes significantly to the sequence error rate. It is noteworthy that modifying the dsDNA template with a hairpin-like adaptor allows obtaining bidirectional (2D) sequencing reads. Like Illuminas paired-end reads, this can quadratically reduce sequencing error rates. Despite these innovations, the accuracy of even 2D reads is significantly worse than even PacBio SMRT sequencing

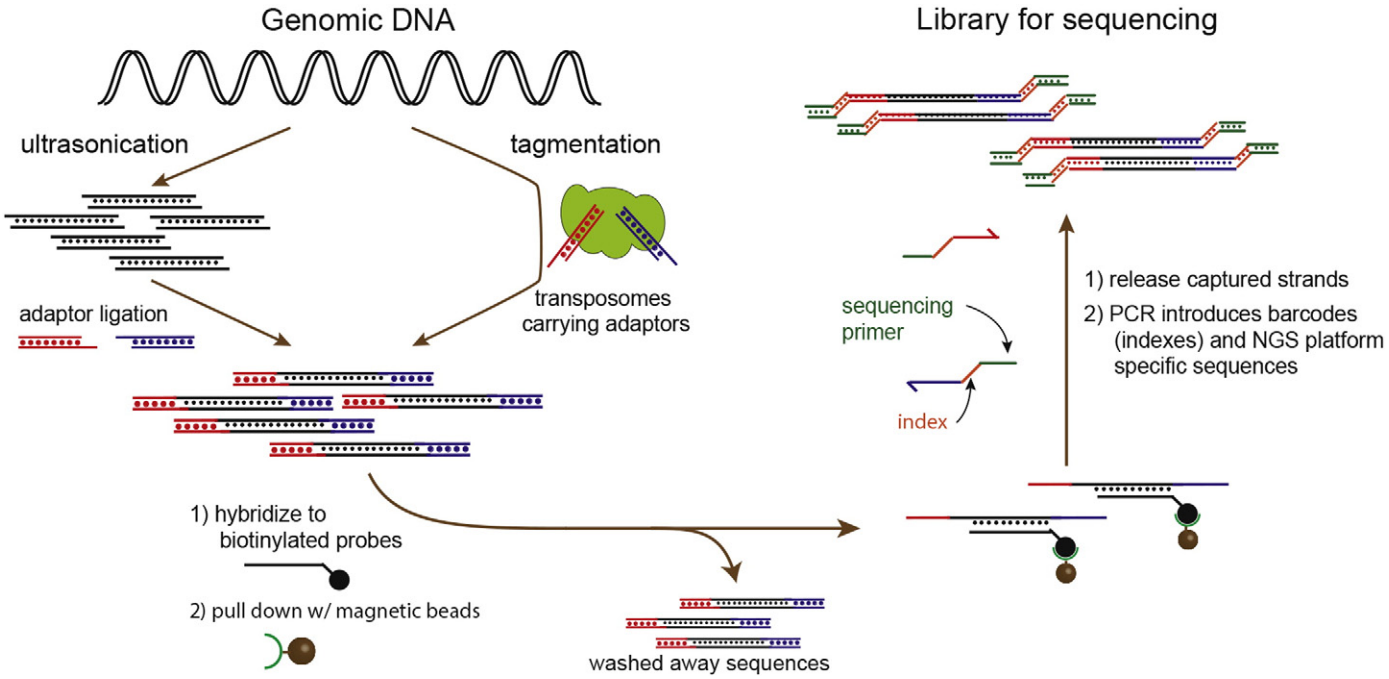

Fig. 10. Hybrid capture target enrichment. Genomic DNA is first sheared physically or enzymatically, and ligated to sequencing adaptors. Biotinylated probes hybridize to desired target sequences; DNA not hybridized to probes are not captured by magnetic beads and washed away. Subsequently, the enriched DNA sample is PCR amplified to introduce NGS specific indices and primers.

[108–111]. Sequencing cost is also high, at $750 per cell, which gives only roughly 40,000 reads. Thus, Oxford Nanopore has not gained significant market adoption.

4.2.3. Genia

Genia Technologies (now part of Roche) also developed a nanopore-based NGS platform. Unlike Oxford Nanopore, Genia sequencing is based on a DNA polymerase tethered to the pore complex. The dNTPs used are modified with polyethylene glycol-based nano-tags of distinct sizes for the 4 different nucleotides (Fig. 9c). As a primer is extended off the target DNA template, the tag is cleaved and flows through the nanopore, inducing an electric current change [112]. Thus far, Genia sequencing has not yet publicly released any sequence data for biological samples, but given Roches $350 M investment in acquiring Genia, it seems likely that a Genia NGS platform will become available within the next couple years.

4.3. Sequence enrichment

The human genome is more than 3 billion nucleotides, and of this the exome comprises roughly 30 million nucleotides (1%). Within the exome, the genes that are related to a particular disease comprise only a small fraction; for example, the Foundation One panel [113] targets roughly 1 million nucleotides across 315 genes, and the Guardant Health 360 panel [114] targets roughly 150 thousand nucleotides across 68 genes. In order to economically obtain sequence information on only target genes of interest, it is necessary to enrich these targets from biological nucleic acid samples.

There are two main approaches to target enrichment: hybrid-capture and multiplexed amplification. Hybrid-capture uses oligonucleotide nucleotide probes on magnetic beads to selectively capture genes/regions of interest, washing away irrelevant sequences. Multiplexed amplification selectively amplifies the genes/regions of interest, increasing their concentration relative to irrelevant sequences. Illumina sequencing has traditionally focused on hybrid-capture, while multiplexed PCR amplification (Ampliseq) is a major attraction of Ion Torrent.

4.3.1. Hybrid capture

The main advantage of hybrid-capture enrichment is its scalability (up to whole exome) and relative lack of bias, and the main disadvantage is the relatively large input DNA requirement (typically 1 μg) and relative slow protocol. The hybrid-capture technique was first introduced in 1991 [115,116], but became popularized for NGS use in 2006 through the development of Agilent's SureSelect technology (Fig. 10) [117].

Hybrid capture target enrichment usually starts with shearing a genomic DNA sample using ultra-sonication, followed by ligation of sequencing adaptors. In contrast, the Illumina Nextera kits [118,119] use transposon complexes to fragment and ligate DNA adapter sequences in one step. Subsequently, the genomic regions with ligated adaptors are PCR amplified both to increase concentration and to introduce indices used for multiplexing different samples within the same NGS run.

At this point, the amplicons are exposed to the biotinylated hybrid-capture probes and hybridization is allowed to occur for 16 to 72 h. Amplicons hybridized to the probes are then captured with streptavidin-coated magnetic beads, while unbound amplicons are washed away. Bound amplicons are subsequently eluted from the beads, typically using sodium hydroxide and/or elevated temperature.

In a comprehensive performance comparison of four major exome enrichment systems [120], Agilent SureSelect showed highest coverage of the intended targets (99.8%), followed by Illumina Nextera, Illumina TruSeq, and Nimble SeqCap EZ (98.2%, 96.9%, and 96.5%, respectively). Agilent SureSelect kit also was shown to be the best for single nucleotide variation (SNV) detection, though Nextera excelled for GC-rich targets (>60% GC). However, all four enrichment systems seemed to struggle with insertion and deletion variations.

Boreal Genomics offers a uniquely different approach to hybrid-capture enrichment, based on differential electrophoretic separation of DNA molecules over a probe-functionalized hydrogel, in the presence of an oscillating electric field at the temperatures close to the melting temperature of the hybrid capture probes [121,122]. Unlike other hybrid-capture techniques that seek to capture a particular set of genes or amplicons, Boreal OnTarget strives to capture specific variants. Proof-of-concept demonstrations show successful enrichment of 46

mutations in 4 genes (Kras, BRAF, EGFR and PIC3CA) in circulating tumor DNA [123]. However, the scalability of this approach appears to be limited by the fact that all probes must be designed to have nearly identical melting temperatures, and by the difficulties common to microarrays (e.g. probe density).

4.3.2. AmpliSeq

Ampliseq is Ion Torrent's flagship enrichment product [124,125] (Fig. 11a). Its main advantages are low input DNA requirements (1 to 10 ng) and fast protocol (3 h), while its main disadvantage is possible primer-dimer artifacts that waste sequencing reads. The Ampliseq workflow can be abstracted as a three-step protocol: target-specific multiplex PCR amplification, primer digestion, and adapter ligation. The first step is guided by Ion Torrent's bioinformatics knowledge base, and the second step is a proprietary enzymatic reaction (Fig. 11a). Ampliseq has achieved remarkable multiplex PCR capability that is currently unmatched; for example, its Ampliseq Comprehensive Cancer panel uses 16,000 primer pairs across 4 tubes, averaging to roughly 4000-plex PCR in each tube [126].

Ion Torrent has developed three translational AmpliSeq-based NGS panels: the Oncomine Comprehensive Assay, the Oncomine Focus Assay, and the Oncomine Cancer Research Panel [127]. These are focused on allowing deep sequencing of cancer related genes to detect rare sequence variants including SNVs, indels, CNVs, and gene fusions.

4.3.3. Droplet-based enrichment

Whereas Ion Ampliseq takes a bioinformatic approach to primer design with clean-up of primer-dimers after amplification, droplet-based enrichment uses the same technology of digital PCR to enable single-plex PCR amplification within each droplet (Fig. 11b). Different droplets contain different primers; thus, many different target sequences may be amplified without dealing with the combinatorial explosion of primer dimer possibilities. Additionally, the primer pair compartmentalization facilitates panel expansion because new primer pairs can be modularly added, in principle without disruption to the existing panel.

Two companies have been primarily working on droplet-based PCR enrichment for NGS: Fluidigm and RainDance. The Fluidigm Access Array system employs parallel amplification of up to 48 different samples with 48 primer pairs, resulting in 2304 individual PCR reactions each 35 nL in volume [152,153]. RainDance Thunderstorm creates millions of picoliter-size droplets, and in principle affords significantly higher multiplexing [154–156]. In practice, both Fluidigm and RainDance multiplexing capabilities lag significantly behind Ampliseq: for example, RainDance's Thunderbolt RUO cancer panel analyzes 230 amplicons [157].

4.3.4. Ligation-based enrichment

Illumina and Agilent developed the ligase-based enrichment methods to enable NGS analysis of low-volume samples (Fig. 12). In both cases, the ligation requires correct binding of separate complementary regions, helping to suppress the impact of nonspecific amplification on NGS analysis.

Illumina TruSeq uses a pair of primers but unlike PCR, both primers bind to the same strand of the template (Fig. 12a) [128]. Extension of the first primer by a non-strand displacing polymerase terminates at the beginning of the second primer; subsequent ligation connects the two primers and their attached adaptors. Nonspecific binding and/or extension of either the first primer or the second primer to other regions of the genome does not result in amplicons with both adaptor sequences. TruSeq reduces the minimum DNA sample size to roughly 50 ng, and allows 1400-plex amplification. The only two FDA cleared NGS assays (for cystic fibrosis) are both based on TruSeq amplicon library enrichment [91,92]. A number of other TruSeq panels are available, including for HLA typing, inherited disease profiling, and autism screening, which may be pursued in the future for clinical IVD use.

Agilent HaloPlex is based on the hybridization of desired genomic fragments to a partially double-stranded DNA probe with sticky ends that bind to both ends of the target DNA [129,130]. Subsequent ligation to the probe results in a circular DNA product that is resistant to nuclease digestion and also amenable to rolling circle amplification

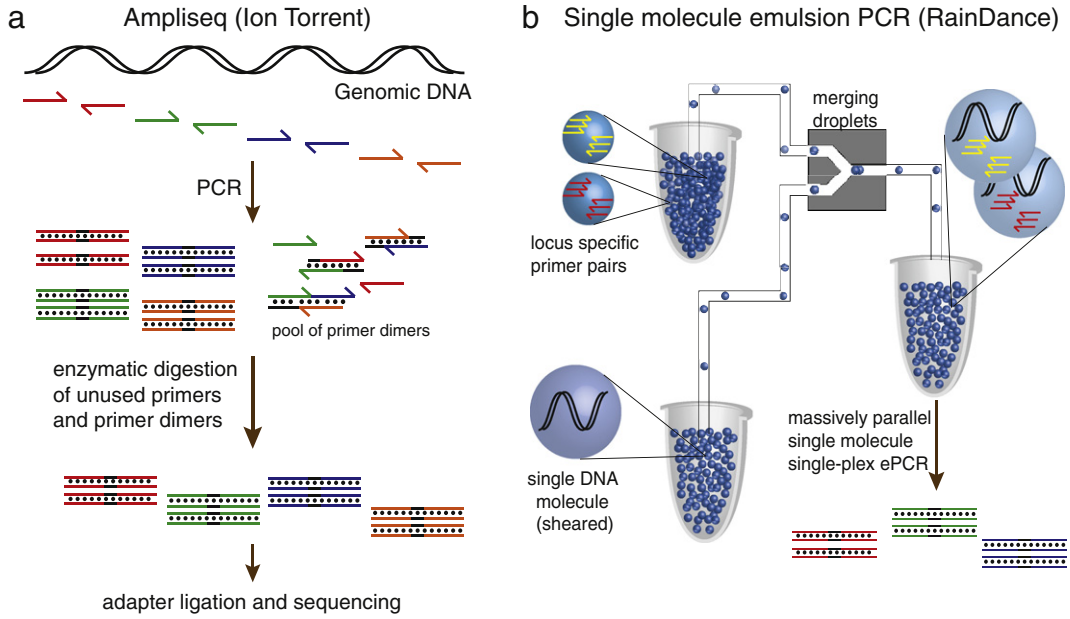

Fig. 11. Multiplex PCR target enrichment. (a) Ion Torrent Ampliseq. Ultra-high multiplex PCR primers are bioinformatically designed, and then the sample is amplified for 8 to 14 cycles. Unused primers and undesired primer dimer amplicons are enzymatically digested. Finally, in order to prepare the obtained amplification products for Ion Torrent sequencing workflow, oligonucleotide adapters consisting of barcodes and sequencing primers are enzymatically ligated. (b) RainDance single molecule emulsion PCR (ePCR). An automated template preparation system emulsifies a sheared DNA sample to produce millions of picoliter-size droplets, each containing one or a few DNA fragments. A microfluidic system then merges each template droplet with another droplet containing a single primer pair specific to a given genetic loci. Thus, each droplet performs a single-plex PCR enrichment.

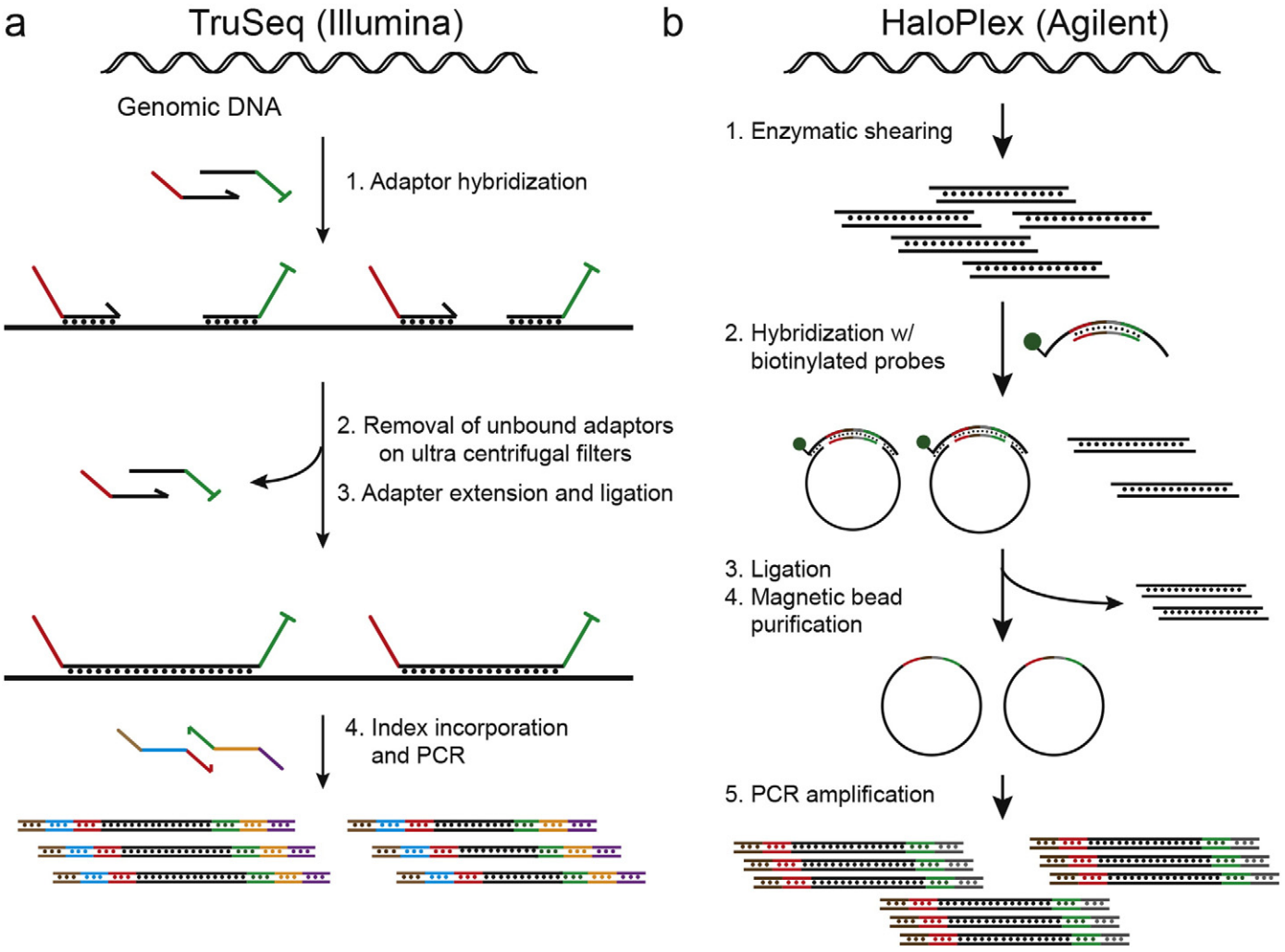

Fig. 12. Ligation-assisted target enrichment. (a) Illumina Truseq uses two oligonucleotide probes that hybridize to regions flanking a target sequence. Polymerase extension of the first probe results in an extended oligonucleotide adjoining the second probe, allowing subsequent ligation. Unreacted probes are removed by centrifugation through size exclusion filters, and successfully ligated products are PCR amplified. (b) Agilent Haloplex hybridizes both ends of a target DNA fragment to a double-stranded probe. Ligation of the target-probe complex produces a circular amplicon that is subsequently purified and amplified.

(Fig. 12b). Currently, Haloplex offers RUO custom designed panels [131, 132] as well as for exome, cancer, and cardiomyopathy gene enrichment.

5. Authors' perspective

Despite the complexity and expense of NGS experiments, NGS is currently the primary choice for many research applications in profiling sequence variation, because it offers unmatched multiplexing throughput and hypothesis-free sequence analysis. Over the coming years, there will likely be two prongs of advance in NGS technology: (1) further increase of throughput for population profiling/screening studies and metagenomics research, and (2) miniaturization of NGS instruments and simplification of NGS workflows for rapid experiments and analysis, both research and clinical. Illumina appears to be the leader (and arguably the only player) in the pursuit of (1), while most other NGS platform developers aim for (2).

The business history of emerging technologies suggests that one to at most two players will emerge victorious in the battle for rapid NGS platform dominance. A number of premature announcements (most notably by Oxford Nanopore in 2012) lacking solid follow-up data have rendered the research community somewhat inured and cynical about novel NGS instruments. To minimize the odds of being discredited as yet another party to cry wolf, a number of NGS instrument developer with functional instruments (e.g. Qiagen GeneReader, BGI Genomics) are likely diligently collecting data in advance of a public announcement.

For clinical diagnostic applications (Box 2), the hypothesis-free nature of NGS is not an advantage, as every reported detection target in a panel must show clinical utility as a requisite for payer reimbursement. Where NGS currently excels is the simultaneous offering of high multiplexing and high allele-sensitivity (microarrays offer high multiplexing and digital PCR offers high allele-sensitivity). In the future, it is likely that additional alternative methods will emerge that combine these features, for diagnostic applications requiring limited multiplexing such as cancer recurrence monitoring. Whether NGS or new approaches win in the clinical setting will depend on a variety of factors including turnaround time, technical performance, cost, pre-analytic sample preparation required, and market penetration/inertia.

The simplicity, robustness, and familiarity of PCR and its variants means that it will likely be the default choice for single-plex nucleic acid analysis for the indefinite future. Digital PCR in particular offers improved quantitation accuracy over traditional quantitative PCR, and may become a staple for clinical diagnostics. Currently, the sensitivity of digital PCR is limited by DNA polymerase nucleotide misincorporation errors, and the detergent formulation of the digital PCR reaction means that it is at

present incompatible with many of the high-fidelity polymerases used for PCR (e.g. Q5, Phusion). This limitation will likely be overcome in the near future through R&D efforts in enzymology or in digital PCR implementation.

5.1. Implications for synthetic biology research

Synthetic biology, as a field, aims to engineer organisms with designed DNA sequences to exhibit engineered behaviors [133,134]. To do so, natural genes must be edited and artificial genes must be introduced to endow organisms with novel functionalities. Purification and verification of error-free genes are important to ensure a homogeneous population of engineered organisms, as even single nucleotide variations can result in loss of gene function [135,136]. To this end, advances in DNA sequence analysis technologies will accelerate the progress of synthetic biology by shortening the design-experiment-analysis cycle, thereby reducing the false-starts from poorly constructed genes. Long-read sequencing (e.g. PacBio), in particular, may be of interest to the synthetic biology community.

RNA sequencing allows highly multiplexed mRNA expression profiling, providing a readout more highly multiplexed than traditional GFP readouts for synthetic biology. This allows easy analysis and debugging of complex engineered circuits and networks as compared to traditional gel electrophoresis or allele-specific PCR approaches. Furthermore, sample barcoding prior to NGS allows simultaneous analysis of many RNA species from many different samples; this is an advantage over microarrays, because each microarray essentially can be used only to analyze one sample. Finally, single-cell sequencing represents the extrapolation limit for NGS sample barcoding, and allows study of expression variability, allowing synthetic biology researchers to observe the stochastic effects of expression in engineered organisms.

5.2. Synthetic biology as a contributor to DNA diagnostics development

Synthetic biology can also contribute to the development of new DNA analysis and diagnostics technologies. In the development of new DNA analysis technologies, well-characterized reference samples are necessary for proof-of-concept testing and validation. CRISPR, the latest generation of DNA editing technologies, uses a guide RNA to sequence-specifically edit a genome at a particular locus [137]. For example, Horizon Discovery uses CRISPR to create cell lines with specific cancer mutations, and mixtures of genomic DNA extracted from these cell lines are popularly used to validate targeted NGS panels.

Synthetic biology has been used to construct bacteria that sense a variety of small molecule targets [138]. Unlike traditional diagnostic devices, these biosensors reproduce, maintaining a population of detectors that can, in principle, indefinitely and continuously report the concentration of target analytes given sufficient nutrients. Given the challenges of immunogenicity, it is unlikely that living biosensors will be used within the next decade as “living diagnostics” for in vivo use in blood. However, synthetic biosensors may find use as low-cost detectors of pathogen DNA and RNA for applications other human diagnostics (e.g. soil profiling and food safety). Although synthetic biological diagnostic devices have thus far focused on small molecule analytes, proof-of-concept RNA detection has recently been demonstrated [139].

6. The future of DNA diagnostics

In the State of the Union address on January 20, 2015, President Obama announced the launch of the Precision Medicine Initiative [140], with the stated goal of bringing “us closer to curing diseases like cancer and diabetes and to give all of us access to the personalized information we need to keep ourselves and our families healthier.” This statement reflects the growing medical and popular understanding that different individuals respond differently to disease and treatment, and that accurate profiling of DNA sequence variations and RNA expression levels are crucial components of the future healthcare paradigm.

In the near term, the focus of research efforts to bring genomics understanding to clinical practice will likely be in the field of cancer-related tests, because cancer is a disease with a plethora of molecular causes, heterogeneity of disease even within an individual, and constant disease evolution. In the longer term, we believe that technology advances will also spill over into diagnostics for all types of human disease, because nearly every disease possess DNA or RNA biomarkers. Infectious diseases diagnostics (including subtyping and antibiotics resistance assays) [141], non-invasive prenatal genetic screening [142,143], and microbiome profiling [144,145], are fields likely to be other early adopters of new DNA technologies.

Technology development will likely play a major role in DNA and RNA diagnostic tests for the foreseeable future. Although NGS throughput and price have dramatically been reduced over the past 10 years (by roughly 100-fold), we remain at least 6 orders of magnitude away from performing comprehensive deep sequencing at a whole genome or whole transcriptome level. Major areas of optimization and innovation will be in the price, accuracy, turnaround time, and multiplexing of tests for DNA sequence variants.

Author contributions

DK, CW, and DYZ wrote the paper. DK and CW contributed equally.

Conflicts of interest

DYZ is a significant equity holder of Searna Technologies. DK, CW, and DYZ have academic collaborations with Thermo Fisher, Nanostring, and DiaCarta.

Acknowledgments

The authors thank Janice Hogan for discussions on FDA and LDT regulations.

This work was funded by the Cancer Prevention Research Institute of Texas (CPRIT) award RP140132 to DYZ.

References

[1] R. Sachidanandam, et al., A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms, Nature 409 (2001) 928–933.
[2] E. Tuzun, et al., Fine-scale structural variation of the human genome, Nat. Genet. 37 (2005) 727–732.
[3] a J. Verkerk, et al., Identification of a gene (FMR-1) containing a CGG repeat coincident with a breakpoint cluster region exhibiting length variation in fragile X syndrome, Cell 65 (1991) 905–914.
[4] B. Kerem, et al., Identification of the cystic fibrosis gene: genetic analysis, Science 245 (1989) 1073–1080.
[5] S. Misale, et al., Emergence of KRAS mutations and acquired resistance to anti-EGFR therapy in colorectal cancer, Nature 486 (2012) 532–536.
[6] M. Peifer, et al., Integrative genome analyses identify key somatic driver mutations of small-cell lung cancer, Nat. Genet. 44 (2012) 1104–1110.
[7] C. Greenman, et al., Patterns of somatic mutation in human cancer genomes, Nature 446 (2007) 153–158.
[8] E. Drenkard, F.M. Ausubel, Pseudomonas biofilm formation and antibiotic resistance are linked to phenotypic variation, Nature 416 (2002) 740–743.
[9] B. Doublet, et al., Variant salmonella genomic island 1-L antibiotic resistance gene cluster in *Salmonella enterica* Serovar Albany, Emerg. Infect. Dis. 9 (2003) 585–591.
[10] Rajyasree Emmadi, et al., Molecular methods and platforms for infectious diseases testing, J. Mol. Diagn. 13 (2011) 583–604.
[11] Z. Zhang, D.W. Chan, The road from discovery to clinical diagnostics: lessons learned from the first FDA-cleared in vitro diagnostic multivariate index assay of proteomic biomarkers, Cancer Epidemiol. Biomark. Prev. 19 (2010) 2995–2999.
[12] S. Marchetti, J.H.M. Schellens, The impact of FDA and EMEA guidelines on drug development in relation to Phase 0 trials, Br. J. Cancer 97 (2007) 577–581.
[13] FDA-approved next-generation sequencing system could expand clinical genomic testing: experts predict MiSeqDx platform will make genetic testing more affordable for smaller labs. AM. J. Med. Genet. A 164, X-XI.
[14] Summary of CAP's Legislative Proposal for the Regulatory Framework of Laboratory-Developed Tests (LDTs), 2015.
[15] Proposal for Modernization of CLIA Regulations for Laboratory Developed Testing Procedures (LDPs), 2015.

[16] Y. Zhao, F. Chen, Q. Li, L. Wang, C. Fan, Isothermal amplification of nucleic acids, Chem. Rev. 115 (22) (2015) 12491–12545.
[17] L.S. Kristensen, L.L. Hansen, PCR-based methods for detecting single-locus DNA methylation biomarkers in cancer diagnostics, prognostics, and response to treatment, Clin. Chem. 55 (2009) 1471–1483.
[18] C.R. Newton, a Graham, L.E. Heptinstall, S.J. Powell, C. Summers, N. Kalsheker, J.C. Smith, a F. Markham, Analysis of any point mutation in DNA. The amplification refractory mutation system (ARMS), Nucleic Acids Res. 17 (1989) 2503–2516.
[19] R.-K. Bai, Detection and quantification of heteroplasmic mutant mitochondrial DNA by real-time amplification refractory mutation system quantitative PCR analysis: a single-step approach, Clin. Chem. 50 (2004) 996–1001.
[20] M. Bergallo, et al., Evaluation of IFN-γ polymorphism + 874 T/a in patients with recurrent tonsillitis by PCR real time mismatch amplification mutation assay (MAMA real time PCR), Cytokine 71 (2015) 278–282.
[21] Barbara Angulo, et al., A comparison of EGFR mutation testing methods in lung carcinoma: direct sequencing, real-time PCR and immunohistochemistry, PLoS One 7 (2012), e43842.
[22] Castro D. Gonzalez de, et al., A comparison of three methods for detecting KRAS mutations in formalin-fixed colorectal cancer specimens, Br. J. Cancer 107 (2012) 345–351.
[23] J. Marchant, A. Mange, M. Larrieux, V. Costes, J. Solassol, Comparative evaluation of the new FDA approved THxIDTM-BRAF test with high resolution melting and sanger sequencing, BMC Cancer 14 (2014) 519.
[24] P.B. McKinzie, B.L. Parsons, Detection of rare K-ras codon 12 mutations using allele-specific competitive blocker PCR, Mutat. Res. 517 (2002) 209–220.
[25] A.Y.C. Tan, D. a Westerman, A. Dobrovic, A simple, rapid, and sensitive method for the detection of the JAK2 V617F mutation, Am. J. Clin. Pathol. 127 (2007) 977–981.
[26] A. Didelot, et al., Competitive allele specific TaqMan PCR for KRAS, BRAF and EGFR mutation detection in clinical formalin fixed paraffin embedded samples, Exp. Mol. Pathol. 92 (2012) 275–280.
[27] Life Technologies. Accurate and Sensitive Somatic Mutation Detection Powered by castPCR Technology, 2014.
[28] C.B. Sherrill, et al., Nucleic acid analysis using an expanded genetic alphabet to quench fluorescence, J. Am. Chem. Soc. 126 (2004) 4550–4556.
[29] K. Reddington, N. Tuite, E. Minogue, T. Barry, A current overview of commercially available nucleic acid diagnostics approaches to detect and identify human gastroenteritis pathogens, Biomol. Detect. Quantif. 1 (2014) 3–7.
[30] http://www.swiftbiosci.com/technology/myt-primers.
[31] Rachel R. Higgins, Melissa Beniprashad, Mark Cardona, Steven Masney, Donald E. Low, Jonathan B. Gubbay, Evaluation and verification of the Seeplex diarrhea-V ACE assay for simultaneous detection of adenovirus, rotavirus, and norovirus genogroups I and II in clinical stool specimens, J. Clin. Microbiol. 49 (2011) 3154–3162.
[32] H. Orum, P.E. Nielsen, M. Egholm, R.H. Berg, O. Buchardt, C. Stanley, Single base pair mutation analysis by PNA directed PCR clamping, Nucleic Acids Res. 21 (1993) 5332–5336.
[33] Y. Nagai, et al., Genetic heterogeneity of the epidermal growth factor receptor in non-small cell lung cancer cell lines revealed by a rapid and sensitive detection system, the peptide nucleic acid-locked nucleic acid PCR clamp, Cancer Res. 65 (2005) 7276–7282.
[34] Diacarta. QClamp KRAS Codon Specific Mutation Detection Kit (Exon 2, 3, 4) Instruction Manual, 2013.
[35] assilios Alexiadis, Tim Watanaskul, Karena Kosco, Lyle Arnold, Abstract: the CEE-selector assay: a tool for the identification of rare allele variants, Cancer Res. 72 (2012) 3198.
[36] J. Li, L. Wang, H. Mamon, M.H. Kulke, R. Berbeco, G.M. Makrigiorgos, Replacing PCR with COLD-PCR enriches variant DNA sequences and redefines the sensitivity of genetic testing, Nat. Med. 14 (2008) 579–584.
[37] C.A. Milbury, J. Li, P.F. Liu, G.M. Makrigiorgos, Cold-PCR: improving the sensitivity of molecular diagnostics assays, Expert. Rev. Mol. Diagn. 11 (2011) 159–169.
[38] G. Santis, et al., Screening for EGFR and KRAS mutations in endobronchial ultrasound derived transbronchial needle aspirates in non-small cell lung cancer using COLD-PCR, PLoS One 6 (2011), e25191.
[39] C.A. Milbury, J. Li, G.M. Makrigiorgos, Ice-COLD-PCR enables rapid amplification and robust enrichment for low-abundance unknown DNA mutations, Nucleic Acids Res. 39 (2011) e2-e2.
[40] J.R. Dobosy, et al., RNase H-dependent PCR (rhPCR): improved specificity and single nucleotide polymorphism detection using blocked cleavable primers, BMC Biotechnol. 11 (2011) 80.
[41] I.M. Mackay, K.E. Arden, A. Nitsche, Real-time PCR in virology, Nucleic Acids Res. 30 (2002) 1292–1305.
[42] R. Molenkamp, A. van der Ham, J. Schinkel, M. Beld, Simultaneous detection of five different DNA targets by real-time Taqman PCR using the Roche LightCycler480: application in viral molecular diagnostics, J. Virol. Methods 141 (2007) 205–211.
[43] S. Tyagi, S.A.E. Marras, F.R. Kramer, Wavelength-shifting molecular beacons, Nat. Biotechnol. 18 (2000) 1191–1196.
[44] M.J. Loeffelholz, D.L. Pong, R.B. Pyles, Y. Xiong, a.L. Miller, K.K. Bufton, T. Chonmaitree, Comparison of the filmarray respiratory panel and prodesse real-time PCR assays for detection of respiratory pathogens, J. Clin. Microbiol. 49 (2011) 4083–4088.
[45] B. Vogelstein, K.W. Kinzler, Digital PCR, Proc. Natl. Acad. Sci. U. S. A. 96 (1999) 9236–9241.
[46] E. Ottesen, J.W. Hong, Stephen R. Quake, Jared R. Leadbetter, Microfluidic digital PCR enables multigene analysis of individual environmental bacteria, Science 314 (2006) 1464–1467.
[47] Valerie Taly, et al., Multiplex picodroplet digital PCR to detect KRAS mutations in circulating DNA from the plasma of colorectal cancer patients, Clin. Chem. 59 (2013) 1722–1731.
[48] N. Wicker, A. Carles, I.G. Mills, M. Wolf, A. Veerakumarasivam, H. Edgren, F. Boileau, B. Wasylyk, J.A. Schalken, D.E. Neal, O. Kallioniemi, O. Poch, A new look towards BAC-based array CGH through a comprehensive comparison with oligo-based array CGH, BMC Genomics 8 (2007) 84.
[49] M. Schena, D. Shalon, R.W. Davis, P.O. Brown, Quantitative monitoring of gene expression patterns with a complementary DNA microarray, Science 270 (1995) 467–470.
[50] C.M. Perou, et al., Molecular portraits of human breast tumors, Nature 406 (2000) 747–752.
[51] T. Yuen, E. Wurmbach, R.L. Pfeffer, B.J. Ebersole, S.C. Sealfon, Accuracy and calibration of commercial oligonucleotide and custom cDNA microarrays, Nucleic Acids Res. 30 (2002), e48.
[52] a Pemov, H. Modi, D.P. Chandler, S. Bavykin, DNA analysis with multiplex microarray-enhanced PCR, Nucleic Acids Res. 33 (2005), e11.
[53] L.B. Baughn, et al., Integration of cytogenomic data for furthering the characterization of pediatric B-cell acute lymphoblastic leukemia: a multi-institution, multi-platform microarray study, Cancer Genet. 208 (2015) 1–18.
[54] Aron C. Eklund, Zoltan Szallasi, Correction of technical bias in clinical microarray data improves concordance with known biological information, Genome Biol. 9 (2008) R26.
[55] R. Salazar, et al., Gene expression signature to improve prognosis prediction of stage II and III colorectal cancer, J. Clin. Oncol. 29 (2011) 17–24.
[56] BenS Wittner, et al., Analysis of the MammaPrint breast cancer assay in a predominantly postmenopausal cohort, Clin. Cancer Res. 14 (2008) 2988–2993.
[57] Christine C. Lee, Gwendolyn A. McMillin, Nikolina Babic, Roberta Melis, Kiang-Teck J. Yeo, Evaluation of a CYP2C19 genotype panel on the GenMark eSensor® platform and the comparison to the Autogenomics InfinitiTM and Luminex CYP2C19 panels, Clin. Chim. Acta 412 (2011) 1133–1137.
[58] R. Pfundt, et al., Clinical performance of the CytoScan Dx assay in diagnosing developmental delay/intellectual disability, Genet. Med. 17 (2015) 1–6.
[59] Kanti Pabbaraju, Sallene Wong, Kara L. Tokaryk, Kevin Fonseca, Steven J. Drews, Comparison of the Luminex xTAG respiratory viral panel with xTAG respiratory viral panel fast for diagnosis of respiratory virus infections, J. Clin. Microbiol. 49 (2011) 1738–1744.
[60] G.K. Geiss, et al., Direct multiplexed measurement of gene expression with color-coded probe pairs, Nat. Biotechnol. 26 (2008) 317–325.
[61] Stephen C. Chapin, David C. Appleyard, Daniel C. Pregibon, Patrick S. Doyle, Rapid microRNA profiling on encoded gel microparticles, Angew. Chem. 123 (2011) 2337–2341.
[62] S.K. Wyman, et al., Post-transcriptional generation of miRNA variants by multiple nucleotidyl transferases contributes to miRNA transcriptome complexity, Genome Res. 21 (2011) 1450–1461.
[63] A.K. Shalek, et al., Single-cell transcriptomics reveals bimodality in expression and splicing in immune cells, Nature 498 (2013) 236–240.
[64] C. Thisse, B. Thisse, High-resolution in situ hybridization to whole-mount zebrafish embryos, Nat. Protoc. 3 (2008) 59–69.
[65] M.J. Gerdes, C.J. Sevinsky, A. Sood, S. Adak, M.O. Bello, Highly multiplexed single-cell analysis of formalin-fixed, paraffin-embedded cancer tissue, Proc. Natl. Acad. Sci. U. S. A. 110 (2013) 11982–11987.
[66] Jeffrey S. Ross, et al., The HER-2 receptor and breast cancer: ten years of targeted anti HER-2 therapy and personalized medicine, Oncologist 14 (2009) 320–368.
[67] Tanner, Minna, et al. Chromogenic in situ hybridization — a practical alternative for fluorescence in situ hybridization to detect HER-2/neu oncogene amplification in archival breast cancer samples. Am. J. Pathol. 157 (5), 2000, 1467–1472.
[68] Y.W. Koh, H.J. Lee, J.W. Lee, J. Kang, G. Gong, Dual-color silver-enhanced in situ hybridization for assessing HER2 gene amplification in breast cancer, Mod. Pathol. 24 (2011) 794–800.
[69] C.-Q. Zhu, et al., Role of KRAS and EGFR as biomarkers of response to erlotinib in National Cancer Institute of Canada Clinical Trials Group Study BR.21. J. Clin. Oncol. 26 (2008) 4268–4275.
[70] E.R. Parilla Castellar, et al., ALK-negative anaplastic large cell lymphoma is a genetically heterogeneous disease with widely disparate clinical outcomes, Blood 124 (2014) 1473–1480.
[71] Tessa M. Grabinski, Andrew Kneynsberg, Fredric P. Manfredsson, Nicholas M. Kanaan, A method for combining RNAscope in situ hybridization with immunohistochemistry in thick free-floating brain sections and primary neuronal cultures, PLoS One 10 (2015), e0120120.
[72] L. Soleymani, Z. Fang, E.H. Sargent, S.O. Kelley, Programming the detection limits of biosensors through controlled nanostructuring, Nat. Nanotechnol. 4 (2009) 844–848.
[73] Z. Fang, L. Soleymani, G. Pampalakis, M. Yoshimoto, J.A. Squire, E.H. Sargent, S.O. Kelley, Direct profiling of cancer biomarkers in tumor tissue using a multiplexed nanostructured microelectrode integrated circuit, ACS Nano 3 (2009) 3207–3213.
[74] H. Lee, T.-J. Yoon, J.-L. Figueiredo, F.K. Swirski, R. Weissleder, Rapid detection and profiling of cancer cells in fine-needle aspirates, Proc. Natl. Acad. Sci. U. S. A. 106 (2009) 12459–12464.
[75] H. Lee, T.-J. Yoon, R. Weissleder, Ultrasensitive detection of bacteria using core–shell nanoparticles and an NMR-filter system, Angew. Chem. Int. Ed. 48 (2009) 5657–5660.
[76] L.P. Samuel, R.J. Tibbetts, A. Agotesku, M. Fey, R. Hensley, F.A. Meier, Evaluation of a microarray-based assay for rapid identification of gram-positive organisms and resistance markers in positive blood cultures, J. Clin. Microbiol. 51 (2013) 1188–1192.

[77] C. Giachetti, et al., Highly sensitive multiplex assay for detection of human immunodeficiency virus type 1 and hepatitis C virus RNA, J. Clin. Microbiol. 40 (2002) 2408–2419.
[78] E.L. Van Dijk, H. Auger, Y. Jaszczyszyn, C. Thermes, Ten years of next-generation sequencing technology, Trends Genet. 30 (2014) 418–426.
[79] M. Eisenstein, The battle for sequencing supremacy, Nat. Biotechnol. 30 (2012) 1023–1026.
[80] R. Squillace, G. Frampton, P.J. Stephens, J.S. Ross, V.A. Miller, Comparing two assays for clinical genomic profiling: the devil is in the data, Onco. Targets. Ther. 8 (2015) 2237–2242.
[81] E. Borgstrom, et al., Phasing of single DNA molecules by massively parallel barcoding, Nat. Commun. 6 (2015) 8173.
[82] D.R. Bentley, et al., Accurate whole human genome sequencing using reversible terminator chemistry, Nature 456 (2008) 53–59.
[83] C. Adessi, et al., Solid phase DNA amplification: characterisation of primer attachment and amplification mechanisms, Nucleic Acids Res. 28 (2000), E87.
[84] M. Fedurco, A. Romieu, S. Williams, I. Lawrence, G. Turcatti, BTA, a novel reagent for DNA attachment on glass and efficient generation of solid-phase amplified DNA colonies, Nucleic Acids Res. 34 (2006), e22.
[85] M.D. Cao, et al., Inferring short tandem repeat variation from paired-end short reads, Nucleic Acids Res. 42 (2014), e16.
[86] J. Zhang, K. Kobert, T. Flouri, A. Stamatakis, PEAR: a fast and accurate Illumina paired-end reAd mergeR, Bioinformatics 30 (2014) 614–620.
[87] A. Bashir, S. Volik, C. Collins, V. Bafna, B.J. Raphael, Evaluation of paired-end sequencing strategies for detection of genome rearrangements in cancer, PLoS Comput. Biol. 4 (2008), e1000051.
[88] S. Lin, Y. Wu, K. Gunderson, J. Moon, Patterned flow-cells useful for nucleic acid analysisUS Patent App. 13/492,661 http://www.google.com/patents/US20120316086 2012.
[89] M.J.R. Shen, et al., Kinetic exclusion amplification of nucleic acid libraries, https://www.google.com/patents/US88952492014.
[90] C. Sheridan, Milestone approval lifts Illuminas NGS from research into clinic, Nat. Biotechnol. 32 (2014) 111–112.
[91] D.S. Grosu, et al., Clinical investigational studies for validation of a next-generation sequencing in vitro diagnostic device for cystic fibrosis testing, Expert. Rev. Mol. Diagn. 14 (2014) 605–622.
[92] M.W. Baker, et al., Improving newborn screening for cystic fibrosis using next-generation sequencing technology: a technical feasibility study, Genet. Med. (2015), http://dx.doi.org/10.1038/gim.2014.209.
[93] B. Merriman, J.M. Rothberg, Progress in Ion Torrent semiconductor chip based sequencing, Electrophoresis 33 (2012) 3397–3417.
[94] M. Stevenson, Thermo raises a new platform for the clinical NGS market, http://www.clpmag.com/2014/09/thermo-raises-new-platform-clinical-ngs-market 2014.
[95] K.F. Au, V. Sebastiano, The transcriptome of human pluripotent stem cells, Curr. Opin. Genet. Dev. 28 (2014) 71–77.
[96] P. Coupland, T. Chandra, M. Quail, W. Reik, H. Swerdlow, Direct sequencing of small genomes on the pacific biosciences RS without library preparation, Biotechniques 53 (2012) 365–372.
[97] L. Cavelier, et al., Clonal distribution of BCR-ABL1 mutations and splice isoforms by single-molecule long-read RNA sequencing, BMC Cancer 15 (2015) 45.
[98] E.W. Loomis, et al., Sequencing the unsequenceable: expanded CGG-repeat alleles of the fragile X gene, Genome Res. 23 (2013) 121–128.
[99] A.E. Minoche, et al., Exploiting single-molecule transcript sequencing for eukaryotic gene prediction, Genome Biol. 16 (2015) 1–13.
[100] J. Eid, et al., Real-time DNA sequencing from single polymerase molecules, Science 323 (2009) 133–138.
[101] S. Howorka, S. Cheley, H. Bayley, Sequence-specific detection of individual DNA strands using engineered nanopores, Nat. Biotechnol. 19 (2001) 636–639.
[102] J. Clarke, et al., Continuous base identification for single-molecule nanopore DNA sequencing, Nat. Nanotechnol. 4 (2009) 265–270.
[103] J.J. Kasianowicz, E. Brandin, D. Branton, D.W. Deamer, Characterization of individual polynucleotide molecules using a membrane channel, Proc. Natl. Acad. Sci. U. S. A. 93 (1996) 13770–13773.
[104] D. Stoddart, G. Maglia, E. Mikhailova, A.J. Heron, H. Bayley, Multiple base-recognition sites in a biological nanopore: two heads are better than one, Angew. Chem. Int. Ed. 49 (2010) 556–559.
[105] A.H. Laszlo, et al., Decoding long nanopore sequencing reads of natural DNA, Nat. Biotechnol. 32 (2014) 829–834.
[106] T.Z. Butler, M. Pavlenok, I.M. Derrington, M. Niederweis, J.H. Gundlach, Single-molecule DNA detection with an engineered MspA protein nanopore, Proc. Natl. Acad. Sci. U. S. A. 105 (2008) 20647–20652.
[107] W. Timp, J. Comer, A. Aksimentiev, DNA base-calling from a nanopore using a viterbi algorithm, Biophys. J. 102 (2012) L37–L39.
[108] A. Kilianski, et al., Bacterial and viral identification and differentiation by amplicon sequencing on the MinION nanopore sequencer, Gigascience 4 (2015) 12.
[109] S. Koren, A.M. Phillippy, One chromosome, one contig: complete microbial genomes from long-read sequencing and assembly, Curr. Opin. Microbiol. 23 (2015) 110–120.
[110] M.-A. Madoui, et al., Genome assembly using nanopore-guided long and error-free DNA reads, BMC Genomics 16 (2015) 327.
[111] E. Karlsson, A. LŠrkeryd, A. Sjšdin, M. Forsman, P. Stenberg, Scaffolding of a bacterial genome using MinION nanopore sequencing, Sci. Rep. 5 (2015) 11996.
[112] S. Kumar, et al., PEG-labeled nucleotides and nanopore detection for single molecule DNA sequencing by synthesis, Sci. Rep. 2 (2012) 1–8.
[113] G.M. Frampton, et al., Development and validation of a clinical cancer genomic profiling test based on massively parallel DNA sequencing, Nat. Biotechnol. 31 (2013) 1023–1031.
[114] 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 10 (2015), e0140712.
[115] M. Lovett, J. Kere, L.M. Hinton, Direct selection: a method for the isolation of cDNAs encoded by large genomic regions, Proc. Natl. Acad. Sci. U. S. A. 88 (1991) 9628–9632.
[116] S. Bashiardes, et al., Direct genomic selection, Nat. Methods 2 (2005) 63–69.
[117] A. Gnirke, et al., Solution hybrid selection with ultra-long oligonucleotides for massively parallel targeted sequencing, Nat. Biotechnol. 27 (2009) 182–189.
[118] A. Adey, et al., In vitro, long-range sequence information for de novo genome assembly via transposase contiguity, Genome Res. 24 (2014) 2041–2049.
[119] S. Picelli, et al., Full-length RNA-seq from single cells using Smart-seq2, Nat. Protoc. 9 (2014) 171–181.
[120] C.S.R. Chilamakuri, et al., Performance comparison of four exome capture systems for deep sequencing, BMC Bioinf. 15 (2014) 449.
[121] J. Pel, et al., Nonlinear electrophoretic response yields a unique parameter for separation of biomolecules, Proc. Natl. Acad. Sci. U. S. A. 106 (2009) 14796–14801.
[122] A. Marziali, J. Pel, D. Bizzotto, L.A. Whitehead, Novel electrophoresis mechanism based on synchronous alternating drag perturbation, Electrophoresis 26 (2005) 82–90.
[123] E. Kidess, et al., Mutation profiling of tumor DNA from plasma and tumor tissue of colorectal cancer patients with a novel, high-sensitivity multiplexed mutation detection platform, Oncotarget 6 (2015) 2549–2561.
[124] D.H. Hovelson, et al., Development and validation of a scalable next-generation sequencing system for assessing relevant somatic variants in solid tumors, Neoplasia 17 (2015) 385–399.
[125] F. Belardinilli, et al., Validation of the Ion Torrent PGM sequencing for the prospective routine molecular diagnostic of colorectal cancer, Clin. Biochem. 48 (2015) 71–73.
[126] Life Technologies, Ion AmpliSeq TM Designer provides full flexibility to sequence genes of your choice, Appl. Note (2013).
[127] Ion Torrent. Oncomine Comprehensive Assay. (2014). at http://www.thermofisher.com/us/en/home/clinical/preclinical-companion-diagnostic-development/oncomine-cancer-research-panel-workflow.html Ion Torrent. Oncomine Focus Assay. (2014). at http://www.thermofisher.com/us/en/home/clinical/preclinical-companion-diagnostic-development/oncomine-focus-assay.html
[128] Illumina, TruSeq Custom Amplicon Library Preparation Guide. 15027983 R, 2013.
[129] Agilent Technologies, HaloPlex Target Enrichment System for Illumina Sequencing, 2015.
[130] F. Miya, et al., A combination of targeted enrichment methodologies for whole-exome sequencing reveals novel pathogenic mutations, Sci. Rep. 5 (2015) 9331.
[131] E.C. Berglund, et al., Accurate detection of subclonal single nucleotide variants in whole genome amplified and pooled cancer samples using HaloPlex target enrichment, BMC Genomics 14 (2013) 856.
[132] L.N.J. Moens, et al., HaloPlex targeted resequencing for mutation detection in clinical formalin-fixed, paraffin-embedded tumor samples, J. Mol. Diagn. 17 (2015) 729–739.
[133] D.E. Cameron, C.J. Bashor, J.J. Collins, A brief history of synthetic biology, Nat. Rev. Microbiol. 12 (5) (2014) 381–390.
[134] D.G. Gibson, et al., Creation of a bacterial cell controlled by a chemically synthesized genome, Science 329 (5987) (2010) 52–56.
[135] A.Y. Borovkov, et al., High-quality gene assembly directly from unpurified mixtures of microarray-synthesized oligonucleotides, Nucleic Acids Res. 38 (19) (2010) e180-e180.
[136] S. Kosuri, G.M. Church, Large-scale de novo DNA synthesis: technologies and applications, Nat. Methods 11 (5) (2014) 499–507 (Chicago).
[137] P. Mali, et al., RNA-guided human genome engineering via Cas9, Science 339 (6121) (2013) 823–826.
[138] J.W. Kotula, S.J. Kerns, L.A. Shaket, L. Siraj, J.J. Collins, J.C. Way, P.A. Silver, Programmable bacteria detect and record an environmental signal in the mammalian gut, Proc. Natl. Acad. Sci. 111 (13) (2014) 4838–4843.
[139] K. Pardee, A.A. Green, T. Ferrante, D.E. Cameron, A. DaleyKeyser, P. Yin, J.J. Collins, Paper-based synthetic gene networks, Cell 159 (4) (2014) 940–954.
[140] Francis S. Collins, Harold Varmus, A new initiative on precision medicine, N. Engl. J. Med. 372 (2015) 793–795.
[141] Eileen M. Burd, Validation of laboratory-developed molecular assays for infectious diseases, Clin. Microbiol. Rev. 23 (2010) 550–576.
[142] H.C. Fan, W. Gu, J. Wang, Y.J. Blumenfeld, Y.Y. El-Sayed, S.R. Quake, Non-invasive prenatal measurement of the fetal genome, Nature 487 (2012) 320–324.
[143] C.F. Wright, H. Burton, The use of cell-free fetal nucleic acids in maternal blood for non-invasive prenatal diagnosis, Hum. Reprod. Update 15 (2009) 139–151.
[144] M.J. Blaser, Harnessing the power of the human microbiome, Proc. Natl. Acad. Sci. U. S. A. 107 (2010) 6125–6126.
[145] P.J. Turnbaugh, R.E. Ley, M. Hamady, C.M. Fraser-Liggett, R. Knight, J.I. Gordon, The human microbiome project, Nature 449 (2007) 804–810.
[146] http://www.fda.gov/MedicalDevices/DeviceRegulationandGuidance/IVDRegulatoryAssistance/ucm123682.htm.
[147] A. Bonington, et al., Use of Roche AMPLICOR mycobacterium tuberculosis PCR in early diagnosis of tuberculous meningitis, J. Clin. Microbiol. 36 (1998) 1251–1254.
[148] A.N. Abou Tayoun, et al., A comprehensive assay for CFTR mutational analysis using next-generation sequencing, Clin. Chem. 59 (2013) 1481–1488.
[149] https://www.cms.gov/Regulations-and-Guidance/Legislation/CLIA/index.html-redirect=/clia/.
[150] M. Cronin, et al., Analytical validation of the Oncotype DX genomic diagnostic test for recurrence prognosis and therapeutic response prediction in node-negative, estrogen receptor-positive breast cancer, Clin. Chem. 53 (2007) 1084–1091.

[151] AdvaMedDx, Introduction to Molecular Diagnostics, 2013 1–19.
[152] A. Kwong, et al., The importance of analysis of long range rearrangement of BRCA1 and BRCA2 in genetic diagnosis of familial breast cancer, Cancer Genet. 208 (2015) 448–454.
[153] V. Lange, et al., Cost-efficient high-throughput HLA typing by MiSeq amplicon sequencing, BMC Genomics 15 (2014) 63.
[154] R. Tewhey, et al., Microdroplet-based PCR enrichment for large-scale targeted sequencing, Nat. Biotechnol. 27 (2009) 1025–1031.
[155] K.E. Ocwieja, et al., Dynamic regulation of HIV-1 mRNA populations analyzed by single-molecule enrichment and long-read sequencing, Nucleic Acids Res 40 (2012) 10345–10355.
[156] N. Tung, et al., Frequency of mutations in individuals with breast cancer referred for BRCA 1 and BRCA 2 testing using next-generation sequencing with a 25-gene panel, Cancer 121 (2015) 25–33.
[157] RainDance, RainDance ThunderBolts TM Cancer Panel, 2015.