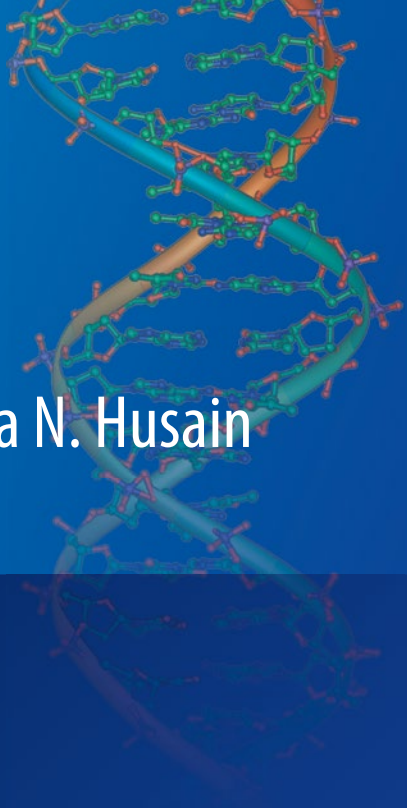
Molecular Pathology Library
Series Editor: Philip T. Cagle



Larissa V. Furtado · Aliya N. Husain
Editors

Precision Molecular Pathology of Neoplastic Pediatric Diseases

Springer

Molecular Pathology Library

Series Editor

Philip T. Cagle
Houston, TX, USA

Creating a diagnostic, prognostic, and therapeutic treatment strategy precisely tailored to each patient's requirements is the fundamental idea behind precision medicine. The Molecular Pathology Library series integrates molecular biology with clinical data for treatment designed for the patient's individual genetic makeup. This approach is widely recognized as the future of medicine and it is vital for practicing pathologists to know the molecular biology, diagnostics and predictive biomarkers for specific forms of cancer or other diseases and their implications for treatment. Each volume focuses on a specific type of cancer or disease and provides concise essential information in a readily accessible, user friendly, convenient format. Each volume is oriented towards the pathologist who needs this information for daily practice, tumor boards, and conferences or for preparation for certification boards or other tests. Written by experts focusing on patient care, these books are indispensible aids to pathologists' participation in precision medicine in the 21st century.

More information about this series at http://www.springer.com/series/7723

Larissa V. Furtado • Aliya N. Husain
Editors

Precision Molecular Pathology of Neoplastic Pediatric Diseases

Springer

Editors
Larissa V. Furtado
Department of Pathology
University of Utah/ARUP Laboratories
Salt Lake City, UT
USA

Aliya N. Husain
Department of Pathology, University of Chicago
Chicago, IL
USA

ISSN 1935-987X ISSN 1935-9888 (electronic)
Molecular Pathology Library
ISBN 978-3-319-89625-0 ISBN 978-3-319-89626-7 (eBook)
https://doi.org/10.1007/978-3-319-89626-7

Library of Congress Control Number: 2018945111

This Springer imprint is published by Springer Nature, under the registered company Springer International Publishing AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Contents

Contributors

John Choi, MD, PhD Department of Pathology, St. Jude Children's Research Hospital, Memphis, TN, USA

Nicole A. Cipriani, MD Department of Pathology, The University of Chicago Medicine and Biological Sciences, Chicago, IL, USA

Carrie Fitzpatrick, PhD Department of Pathology, University of Chicago Medical Center, Chicago, IL, USA

Larissa V. Furtado, MD Department of Pathology, University of Utah/ARUP Laboratories, Salt Lake City, UT, USA

M. John Hicks, MD, DDS, MS, PhD Department of Pathology and Immunology, Texas Children's Hospital and Baylor College of Medicine, Houston, TX, USA

Department of Pediatrics, Division of Hematology and Oncology, Texas Children's Hospital and Baylor College of Medicine, Houston, TX, USA

Department of Pediatric Dentistry, School of Dentistry, University of Texas Health Science Center at Houston, Houston, TX, USA

Aliya N. Husain, MD Department of Pathology, University of Chicago, Chicago, IL, USA

Naohiko Ikegaki, PhD Department of Anatomy and Cell Biology, University of Illinois at Chicago, Chicago, IL, USA

Jason A. Jarzembowski, MD, PhD Division of Pediatric Pathology, Medical College of Wisconsin, Milwaukee, WI, USA

Nasim Khoshnam Georgia Institute of Technology and Emory University School of Medicine, Atlanta, GA, USA

Jeffery M. Klco, MD, PhD Department of Pathology, St. Jude Children's Research Hospital, Memphis, TN, USA

Selene C. Koo, MD, PhD Department of Pathology and Laboratory Medicine, Nationwide Children's Hospital, Columbus, OH, USA

Thomas Krausz, MD, FRCPath University of Chicago Medicine, Chicago, IL, USA

Kamran Mirza, MD, PhD Department of Pathology and Laboratory Medicine, Loyola University Chicago Stritch School of Medicine, Maywood, IL, USA

Kristen M. Paral, MD Duke University Medical Center, Durham, NC, USA

Bruce R. Pawel, MD Division of Anatomic Pathology, Department of Pathology and Laboratory Medicine, The Children's Hospital of Philadelphia and The Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA

Elizabeth J. Perlman, MD The Ann & Robert H. Lurie Children's Hospital of Chicago, Chicago, IL, USA

Navin R. Pinto, MD Department of Pediatrics, Seattle Children's Hospital, Seattle, WA, USA

Peter Pytel, MD Department of Pathology, The University of Chicago Medicine and Biological Sciences, Chicago, IL, USA

Sarangarajan Ranganathan, MD Department of Pathology, Children's Hospital of Pittsburgh of UPMC, Pittsburgh, PA, USA

Lauren Ritterhouse, MD, PhD University of Chicago, Chicago, IL, USA

Jeremy P. Segal, MD, PhD University of Chicago, Chicago, IL, USA

Bahig M. Shehata, MD Georgia Institute of Technology and Emory University School of Medicine, Atlanta, GA, USA

Hiroyuki Shimada, MD, PhD Department of Pathology and Laboratory Medicine, Children's Hospital Los Angeles, University of Southern California, Keck School of Medicine, Los Angeles, CA, USA

Elizabeth A. Sokol, MD Department of Pediatrics, University of Chicago Comer Children's Hospital, Chicago, IL, USA

Rachel L. Stewart, DO, PhD Department of Pathology and Laboratory Medicine, University of Kentucky, Lexington, KY, USA

Part I
Molecular Pathology of Pediatric Neoplasms: General Principles

Chapter 1
Principles of Molecular Biology and Oncogenesis

Rachel L. Stewart, Selene C. Koo, and Larissa V. Furtado

Principles of Molecular Biology

Nucleic Acid Structure, Composition, and Organization

Nucleic acids are macromolecules comprised of chains of nucleotides. Each nucleotide consists of a sugar, a phosphate group, and a nitrogenous base. The nitrogenous bases (purines and pyrimidines) are aromatic heterocyclic compounds. The purines (guanine and adenine) consist of two carbon-nitrogen rings, while the pyrimidines (cytosine and thymine) each consist of a single ring. In the double helix formed by deoxyribonucleic acid (DNA), purines pair with pyrimidines through hydrogen bonds. Base pairing of adenine (A) with thymine (T) leads to the formation of two hydrogen bonds, while base pairing of cytosine (C) and guanine (G) leads to the formation of three hydrogen bonds. Hydrogen bonding contributes to the thermodynamic stability of DNA as well as to its unique double helical structure.

DNA is composed of two polynucleotide chains that twist around each other resulting in the formation of the quintessential double helix. Each strand of the double helix contains alternating sugar and phosphate groups that are referred to as the sugar-phosphate backbone. In DNA, the sugar is 2-deoxyribose, a 5-carbon sugar that lacks a

R. L. Stewart, DO, PhD (✉)
Department of Pathology and Laboratory Medicine, University of Kentucky, Lexington, KY, USA
e-mail: rachel.stewart@uky.edu

S. C. Koo, MD, PhD
Department of Pathology and Laboratory Medicine, Nationwide Children's Hospital, Columbus, OH, USA

L. V. Furtado, MD
Department of Pathology, University of Utah/ARUP Laboratories, Salt Lake City, UT, USA

L. V. Furtado, A. N. Husain (eds.), *Precision Molecular Pathology of Neoplastic Pediatric Diseases*, Molecular Pathology Library, https://doi.org/10.1007/978-3-319-89626-7_1

hydroxyl group at the 2′ position. The combination of a five-carbon sugar and a nitrogenous base is called a nucleo*side*, and the addition of a phosphate group results in the formation of a nucleo*tide*. Nucleotides are linked together through the formation of phosphodiester bonds, covalent bonds that join the 5′-phosphate group of one nucleotide to the 3′-hydroxyl group of another. The two chains of the double helix are complementary to each other, and because of base pairing rules, if we know the sequence of nucleotides on one chain, then we can easily deduce the sequence of nucleotides on the opposite chain. For example, if we have the sequence 5′-ATCGCT-3′, then the complementary sequence is 3′-TAGCGA-5′. While the sugar-phosphate backbone in DNA maintains the same repetitive order of alternating sugars and phosphates along its backbone, the unique sequence of nucleotides with differing nitrogenous bases provides an extraordinary method for the storage of biological information.

The central dogma of molecular biology describes the direction of flow for genetic information: DNA serves as a template for the transcription of RNA, and from RNA the synthesis of proteins is directed in a process called translation. Generally speaking, this flow of information is unidirectional: DNA→ RNA→ protein. RNA is a nucleic acid that is very similar to DNA, though it differs in a few key attributes. Whereas the 5-carbon sugar of DNA is 2-deoxyribose, the sugar in RNA is called ribose and contains an additional hydroxyl group on the 2′ carbon. Furthermore, while DNA contains the nitrogenous base thymine, RNA utilizes the closely related base uracil. In RNA, base pairing occurs between adenine and uracil and between cytosine and guanine. RNA typically exists as a single-stranded nucleic acid, although in some situations, RNAs may fold and twist in such a manner as to simulate a double-stranded structure.

Gene Structure, Organization, and Expression

Gene expression involves the process of transcription, during which DNA is transcribed into RNA. Transcription of messenger RNA (mRNA) is catalyzed by the enzyme RNA polymerase II. The initiation of transcription begins when RNA polymerase II (and associated molecules) assembles at a specific site in DNA called the promoter. There, RNA polymerase II separates the strands of the DNA helix so that it can direct the synthesis of mRNA by using one DNA strand as a template. RNA polymerase II then adds complementary nucleotides sequentially, thus elongating the newly formed RNA molecule. This process continues until a polyadenylation signal is reached, at which point the transcription process is terminated. mRNAs then undergo posttranscriptional modifications, including polyadenylation, capping, and splicing. It is important to note that during transcription, both the coding (exons) and noncoding (introns) regions of DNA are transcribed, so that newly synthesized RNA contains large regions of intervening genetic material that does not encode for amino acids. These intervening regions are removed through a process known as splicing. Splicing generally requires the presence of certain consensus sequences immediately upstream or downstream of the exon-intron junction. In alternative splicing, the products of genes may be modified through the inclusion or exclusion of exons in the final processed mRNA.

The genetic information contained within the final processed mRNA is translated into protein through the use of a specialized genetic code. In this code, nucleotide sequences found in mRNA are translated into amino acids. Each amino acid is encoded by a nucleotide triplet that is referred to as a *codon*. There are 64 possible codons, and each of these codons specifies for one of 20 unique amino acids. For example, the codon GCC codes for the amino acid *alanine*, while the codon GGC codes for *glycine*. Codons can also specify translation start sites (AUG—*methionine*) and termination sites (e.g., UAG—*stop*).

Individual genes contain protein-coding regions called exons. Exons are the sequence regions that remain after mRNA splicing and intron removal. As mentioned previously, alternative splicing can result in the production of different proteins from the same gene sequence. The number of exons in a gene can vary from as few as 1 to >150. The nucleotide sequences within exons are translated into proteins using the genetic code described above.

The human genome contains approximately three billion base pairs and is estimated to contain at least 20,000 genes. Although the amount of information in the human genome is vast, not all of it directly codes for functional proteins. In fact, the majority of the DNA in our genomes consists of intervening sequences and introns, also referred to as noncoding DNA. In previous years, these regions of untranslated sequence were thought to be biologically unimportant; however, it is now apparent that these regions have a variety of biologically relevant functions. Noncoding DNA specifies for a diverse array of RNA subtypes, including ribosomal, transfer, and microRNAs. MicroRNAs are important for gene silencing and for regulating gene expression. Regulatory elements, pseudogenes, and telomeres are just a few other examples of functionally important noncoding DNA sequences.

Chromosome Structure

The majority of human DNA is contained within the cell nucleus, where it is divided into organizational units called chromosomes. In eukaryotic cells, a chromosome consists of a single, extremely long DNA molecule along with associated proteins that are involved in packaging and storing DNA. The majority of human cells are diploid, meaning that each cell has two copies of each chromosome (22 pairs of autosomes and 1 pair of sex chromosomes) for a total of 46 chromosomes. These chromosomes contain long stretches of genes, as well as even longer stretches of noncoding DNA. If laid out longitudinally, this vast amount of genetic material would be nearly 1 m long per cell. In order to pack this massive amount of genetic material into the nucleus of an individual cell, DNA associates with histone proteins to form structures termed *nucleosomes*. DNA wraps around histone proteins in a manner analogous to thread wrapping around a spool. This is also referred to as having the appearance of "beads on a string." In this way, the linear length of DNA can be compressed to fit within a single cell. One functional consequence of this process is that it modifies DNA accessibility and has functional consequences for gene expression.

Mitochondrial DNA

In humans, the majority of genetic information is stored within chromosomes in the cell nucleus; however, a small amount of DNA can also be found in cellular organelles called mitochondria. This DNA is referred to as *mitochondrial DNA* or *mtDNA*. A major function of mitochondria is to generate ATP through a process known as oxidative phosphorylation, and correspondingly, mtDNA encodes a number of enzymes that are required for oxidative phosphorylation. The mitochondrial genome is quite small when compared to that of nuclear DNA and consists of only ~16,000 base pairs. Mitochondrial DNA is unique in that it is packaged into a double-stranded, circular genome. One of the DNA strands is referred to as the *heavy strand* and is rich in guanine, while the other is referred to as the *light* strand and is rich in cytosine. Each mitochondrion has two to ten copies of mitochondrial DNA, and each cell has between 1000 and 2000 mitochondria. Thirty-seven genes are present in mitochondrial DNA, and many of these genes encode for enzymes that are involved in oxidative phosphorylation. In addition to genes encoding metabolic enzymes, mitochondrial DNA also encodes various ribosomal RNAs (rRNAs) and transfer RNAs (tRNAs).

Mitochondrial DNA is inherited maternally, meaning the mtDNA sequences between maternally related individuals are identical in the variable regions (as long as there are no mutations present). Mitochondrial DNA is clinically important for a number of reasons. Mitochondrial disorders, also known as oxidative phosphorylation disorders, can result from mutations or genomic alterations in mitochondrial DNA. These disorders tend to affect organs that are highly dependent on oxidative phosphorylation, including the heart, skeletal muscle, brain, and kidney. Disorders involving mitochondrial genes include Leber hereditary optic neuropathy (LHON) and mitochondrial encephalomyopathy, lactic acidosis, and stroke-like episodes (MELAS). MtDNA testing is used in the diagnosis and characterization of mitochondrial disorders and is also used in forensic science.

Principles of Oncogenesis

Cancer is a genetic disease. The classical understanding of cancer is that there is an inciting event that results in dysregulation of normal cell growth, leading to inappropriate growth and differentiation. Most cancers are sporadic and result from genetic alterations in somatic cells, although some cancers are inherited, with genetic alterations in germline cells conferring increased susceptibility to cancer development. The process of oncogenesis is complex, dynamic, and often multistep. It is likely dependent on the acquisition of several biological capabilities by somatic cells that result in their independence to external growth signals, insensitivity to external anti-growth signals, indefinite replication, evasion of apoptosis, sustained angiogenesis, activation of tissue invasion and metastasis, reprogramming of energy metabolism, and evasion of host immune response [1, 2]. These oncogenic hallmarks are enabled by genome instability, mutations, and tumor-promoting inflammation [1]. Cancer results from coordinated and complementary

functional changes in multiple pathways, and its development requires feedback interactions between cancer cells and their microenvironment, composed of a repertoire of non-neoplastic cells within a systemic context involving inflammation, immune responses, and metabolism [3]. However, it is important to consider that cancer is not a single disease and that the consequences of using shared molecular pathways vary among different tumor types [4]. In addition, the different rates of stem cell division among different tissues contribute to the variation in cancer risk among different tissues [5].

The most frequent cancer-associated alterations are point mutations (single-nucleotide substitutions), insertions, deletions, duplications, gene amplification, gene rearrangements, and copy number variations, which may affect coding regions, splicing sites, or promoters of genes. Coding region mutations are classified, based on their effect on the codon, as silent mutations, when the coded amino acid is not changed by a nucleotide change; missense mutations, in which a nucleotide substitution results in the replacement of an amino acid to another; or nonsense mutations, which replace the coded amino acid with premature termination of protein translation and protein truncation. The genes that are mutated in cancer are broadly divided into two major categories: oncogenes and tumor suppressor genes.

Oncogenes

Oncogenes are mutated forms of cellular proto-oncogenes. Proto-oncogenes are generally genes that are important in regulating cell growth and differentiation. A genetic alteration that results in dysregulated activation results in conversion of the proto-oncogene to an oncogene. Proto-oncogenes generally require only one activating or gain-of-function mutation to become oncogenic. Usual types of mutations that result in proto-oncogene activation include point mutations, gene amplifications, and chromosomal translocations. Typically, the genetic alterations that result in oncogene activation occur within specific codons or clusters of codons or mutational hotspots. For example, a single activating mutation, c.1799 T > A (p.V600E), within the activation segment of the tyrosine kinase domain of the protein kinase *BRAF* results in constitutive kinase activation and insensitivity to negative feedback [6, 7] and drives tumorigenesis in the majority of papillary thyroid carcinomas and ameloblastomas [8, 9]. Another example is mutations in the tyrosine kinase (TK) domain of the epidermal growth factor receptor (*EGFR*) gene that occur in non-small cell lung cancer, which lead to constitutive activation of EGFR kinase activity and downstream signaling.

Tumor Suppressor Genes

Tumor suppressor genes encode proteins that inhibit cell proliferation and block the development of tumors. Since a single functional copy of a gene is typically sufficient for adequate protein function within the cell, tumor suppressor genes

classically follow the two-hit model of tumorigenesis, in which both alleles of a gene must be mutated before a tumor can develop. Many cancer predisposition syndromes are caused by germline mutations in a single allele of a tumor suppressor gene, providing a first hit that requires only a single additional somatic inactivating hit, either through point mutations, deletions, or epigenetic gene inactivation, to trigger tumor formation. For example, in approximately 40% of patients with the pediatric eye tumor retinoblastoma, mutation of one copy of the tumor suppressor gene *RB1* is inherited (germline) and found in all cells of the child's body; mutation of the other *RB1* copy in retinal cells leads to development of retinoblastoma. These children often develop retinoblastoma in both eyes. In sporadic retinoblastoma, there is no germline *RB1* mutation, and two separate mutation events in *RB1* are required within a single cell to develop the tumor. Tumor suppressor genes, such as *TP53* and *CDKN2A*, typically encode for proteins that regulate cell cycle checkpoint responses, DNA damage detection and repair, apoptosis, and differentiation.

In addition to the aforementioned molecular alterations, microRNA dysfunction and environmental factors (e.g., carcinogenic chemicals, radiation, and viral or bacterial infections) are also associated with the pathogenesis of certain human cancers via disruption of gene(s) involved in the control of cell growth and division.

Knowledge about oncogenic hallmarks has served as a powerful guide for translational research aimed at developing many areas of personalized cancer care, including screening, diagnosis, therapy selection, therapeutic monitoring, and prognosis through detection or measurement of molecular biomarkers. It is anticipated that future studies on the interactions between tumor cells and their microenvironment may shed further light on factors contributing to the development and progression of cancer, with the hope that these will ultimately provide additional tools for personalized cancer care.

References

1. Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. Cell. 2011;144(5):646–74.
2. Hanahan D, Weinberg RA. The hallmarks of cancer. Cell. 2000;100(1):57–70.
3. Hainaut P, Plymoth A. Targeting the hallmarks of cancer: towards a rational approach to next-generation cancer therapy. Curr Opin Oncol. 2013;25(1):50–1.
4. Fouad YA, Aanei C. Revisiting the hallmarks of cancer. Am J Cancer Res. 2017;7(5):1016–36.
5. Tomasetti C, Vogelstein B. Cancer etiology. Variation in cancer risk among tissues can be explained by the number of stem cell divisions. Science. 2015;347(6217):78–81.
6. Davies H, Bignell GR, Cox C, et al. Mutations of the BRAF gene in human cancer. Nature. 2002;417(6892):949–54.
7. Pratilas CA, Taylor BS, Ye Q, et al. (V600E)BRAF is associated with disabled feedback inhibition of RAF-MEK signaling and elevated transcriptional output of the pathway. Proc Natl Acad Sci U S A. 2009;106(11):4519–24.
8. Henke LE, Perkins SM, Pfeifer JD, et al. BRAF V600E mutational status in pediatric thyroid cancer. Pediatr Blood Cancer. 2014;61(7):1168–72.
9. Kurppa KJ, Caton J, Morgan PR, et al. High frequency of BRAF V600E mutations in ameloblastoma. J Pathol. 2014;232(5):492–8.

Chapter 2
Molecular Methods in Oncology: Targeted Mutational Analysis

Jason A. Jarzembowski

PCR-Based Techniques

Overview

Polymerase chain reaction (PCR) is, at its core, a method of quickly and massively amplifying a particular DNA region of interest so it can be analyzed. PCR uses carefully selected oligonucleotide primers complementary to the ends of the target sequence in a cyclical reaction that recruits the products of each round to serve as additional template in the next round, resulting in exponential yields. The sensitivity and specificity of PCR allow a wide range of source materials to be used, including fresh, frozen, and formalin-fixed tumor tissue. PCR is widely used for mutational analysis, detecting fusion genes, measuring gene expression, and determining gene methylation status.

Background

Polymerase-based amplification of short segments of DNA using oligonucleotide primers was first described in a 1971 paper from Khoruna's laboratory, but further refinement and realization of the method's potential was done over a decade later by Kerry Mullis, who would later receive the Nobel Prize in Chemistry for his work [1, 2]. As first conceived and implemented, PCR required manual movement of the samples between three different water baths and the addition of new enzyme with every cycle. Since then, the development of bench-top thermocyclers and

J. A. Jarzembowski, MD, PhD
Division of Pediatric Pathology, Medical College of Wisconsin, Milwaukee, WI, USA
e-mail: jjarzemb@mcw.edu

L. V. Furtado, A. N. Husain (eds.), *Precision Molecular Pathology of Neoplastic Pediatric Diseases*, Molecular Pathology Library, https://doi.org/10.1007/978-3-319-89626-7_2

thermostable DNA polymerase has greatly facilitated the procedure, and innumerable variations have been created, making PCR an oft-used and indispensable technique in every clinical and research laboratory [3].

Method

Template

PCR is a versatile method that can utilize a variety of template DNA sources. Only very small quantities of DNA (nanograms or less) are required for most applications; however, the purity and quality of the DNA are often critical factors. Controls are needed in every PCR using known amounts of purified DNA with the assay primers, as well as using primers against control ("housekeeping") genes with the patient DNA to ensure that reactions work, reagents are good, and inhibitors are not present.

Many tumors will only have formalin-fixed paraffin-embedded tissue available for use, and this usually suffices [4, 5]. However, formalin fixation creates DNA-DNA and DNA-protein crosslinks that can interfere with PCR; both can sterically hinder the progression of polymerase or interfere with primer-template binding. Longer fixation times allow for more crosslinks and more difficulty with subsequent analysis [6, 7]. Acid decalcification destroys nucleic acid, and such specimens are not usually suitable for PCR (or other molecular testing) [8].

Thus, it is beneficial (though not required) for any specimens on which PCR or other nucleic acid-based assays may be performed to have a tissue aliquot snap-frozen and reserved for this purpose. Colder is better—the DNA in tissue specimens stored at −70 °C is adequately preserved for years to decades, at −20 °C for months to a year, and at 4 °C for perhaps days to weeks [9, 10].

Kits for DNA extraction are commercially available and generally work well. Most include proteinase (to remove proteins) and ribonuclease (RNase) treatment (to remove RNA) followed by organic or solid-phase extraction to purify the DNA. Spectrophotometry can be used to measure the concentration and assess the purity of the DNA, both of which are important for successful PCR. DNA yield may also be quantitated via fluorometric quantitation of double-stranded DNA (e.g., Qubit™, PicoGreen®), via electrophoresis-based methods (e.g., agarose gel followed by ethidium bromide staining, Bioanalyzer, TapeStation), or via real-time PCR (qPCR). The quality/integrity of DNA samples may be assessed using an electrophoresis-based assay, where intact DNA will appear as a high molecular weight single band, while degraded DNA is identified as a smear of variably sized fragments, as well as via qPCR or by using an amplification control in the assay.

RNA is markedly more labile in tissue because of endogenous and exogenous endonucleases and requires snap-freezing samples within the first hour of tissue procurement for reliable and reproducible results. In order to improve RNA preservation, special precautions are necessary, including the use of diethyl pyrocarbonate

(DEPC) water in all reagents used in RNA procedures and decontamination of work area and pipettes to prevent RNase contamination. For RNA-based protocols, extraction protocols are similar but use a deoxyribonuclease instead of a ribonuclease, and the purified RNA is used in a reverse transcriptase reaction to create cDNA, which then serves as the template for PCR.

Primers

The target sequence which will be amplified is defined by the selection of oligonucleotide primers that flank it. The "upstream" or "forward" primer is complementary to the minus strand of the DNA template (recapitulates the sense sequence), and the "downstream" or "reverse" primer is complementary to the plus strand (reads as antisense). Primers are usually 20–40 bases long and are chosen in order to amplify the sequence of interest as well as to be relatively unique in the genome and have a sufficiently high annealing temperature to decrease the synthesis of nonspecific products. The annealing temperature depends on a host of factors, most notably the primer length and its GC content (remember guanine and cytosine pair with three hydrogen bonds, as compared to two bonds for the adenine-thymine pair, so GC-rich sequences melt at a higher temperature). One of the most important characteristics of a primer in determining specificity is the 3′-most sequence; while nonhomologous linkers and point mismatches can be present at or near the 5′ end, the 3′ end must match in order to firmly bind the template and allow polymerase to initiate [11].

Obviously, the main selection criteria for primers will be that they amplify the sequence of interest. Although the target must contain the specific gene or mutation site to be analyzed, there is usually some flexibility in the precise 5′ and 3′ stop and start sites, which allows optimization of primers based on the criteria described above. However, certain limitations of target sequence and length exist. For example, highly repetitive sequences can be difficult to amplify because of the tendency for primer and template to form secondary structures or "slip" and bind to a neighboring site when amplifying repetitive regions [12]. A reasonable estimate is elongation of about 1 kb/min and a maximum length of 5 kb under routine conditions. Optimized reactions can amplify targets of >20 kb, but this is less practical for routine use and often impossible when using DNA extracted from FFPE tissues.

Polymerase

Several different DNA polymerases can be employed in PCR. One common trait they all share is being thermostable, which both allows them to survive the high temperature denaturation cycles (>90 °C) without being denatured themselves, as well as to function at a high enough temperature (70–75 °C) where

nonspecific primer binding does not easily occur. Some of these DNA polymerases are originally from thermophilic bacteria (such as *Taq* polymerase from *Thermus aquaticus*) that live in hot springs and similar environments, some are genetically modified variants of DNA polymerase from other sources, and some are both [13].

In addition to their primary function of elongating a DNA strand in a 5′→3′ direction according to a second strand template, all DNA polymerases also have an intrinsic 5′→3′ exonuclease activity, meaning they are able to remove segments of DNA that they encounter on their way, in the direction they are moving. This allows them to continue elongation without being disrupted, and this function is exploited in real-time PCR (see below). However, one of the major differences between thermostable DNA polymerases used in PCR is whether they also possess a 3′→5′ exonuclease activity. This so-called proofreading function allows the enzyme to remove the previously added nucleotide when it is incorrectly matched to the template. The 3′→5′ exonuclease, which is absent from *Taq* but present in Pfu, Tfu, and Vent polymerases, can reduce the error rate of mismatched bases by about 100-fold, from about 1 in 10,000 bases to 1 in 1,000,000 [13]. The increased fidelity is important in PCR applications such as mutational analysis where the sequence will be scrutinized and where base errors in early rounds of amplification could be propagated and interfere with the final results.

Procedure

PCR amplification involves iterative rounds of denaturation, annealing, and extension, with the products of each cycle serving as additional template in subsequent cycles (Fig. 2.1). A standard reaction contains sequence-specific primers, free deoxynucleotide triphosphates (dNTPs), template DNA, and thermostable DNA polymerase in a buffer. The reaction is heated to 92–95 °C to denature the double-stranded DNA template and then cooled to allow the primers to anneal to the now single-stranded template. The annealing temperature varies depending on the length and sequence (GC content) of the primers but is typically between 50 and 60 °C. The reaction is then warmed to 70–75 °C, the optimal temperature for the polymerase to extend the primers based on their bound template sequence, incorporating the free dNTPs. At the end of this cycle, the primer-template molecules will have been extended into double-stranded DNA, thus (in theory) doubling the concentration of the template for the next round.

Like any process, PCR is not 100% efficient—primer binding is not complete or exact, some template strands will not be fully extended, and reactants will become depleted during the reaction—so the actual amplification will fall somewhat short of exponential. Also, the initial template strands can be elongated indefinitely (as far as time allows, well past the second primer site) in each cycle, whereas the products will be "hemmed" in by the second primer and will have the exact length of the desired product. The former are amplified only linearly and contribute to the sub-

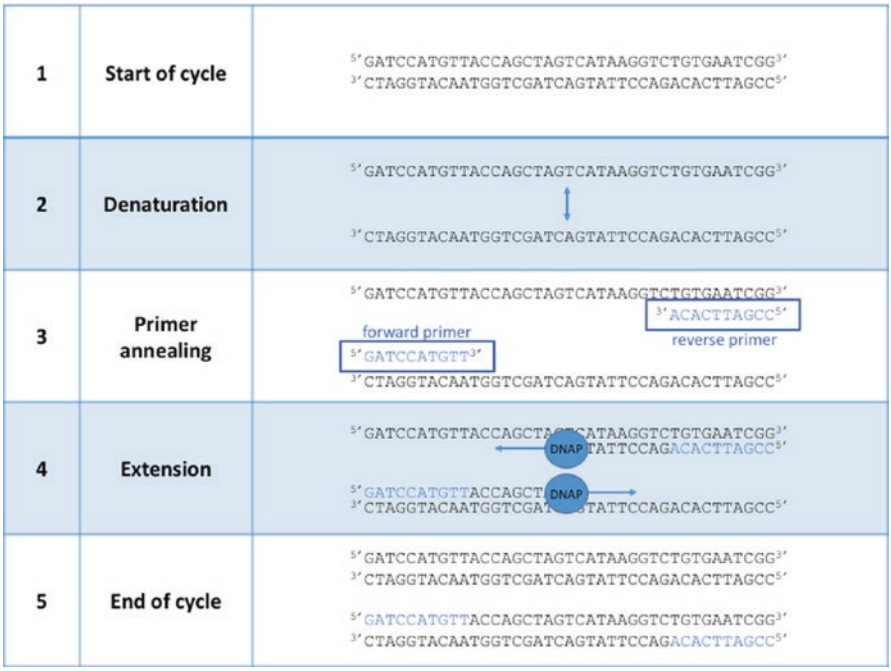


Fig. 2.1 Overview of PCR technique. PCR begins with a double-stranded DNA template (1) that is heat-denatured to separate the strands (2). As the reaction cools to the annealing temperature, sequence-specific forward and reverse primers bind to the ends of the target sequence (3). The reaction is then warmed to an optimal condition for the thermostable DNA polymerase (DNAP), which elongates each of the primers according to the complementary template strand (4). At the end of each cycle, the amount of product will have doubled (theoretically), and the newly created product serves as additional template in the next cycle (5)

theoretical reaction yield. The temperature and duration of each step, as well as the number of cycles (usually 20–30), can have profound effects on the product yield and amount of nonspecific product formed.

At the conclusion of the reaction, the PCR products can be assessed in several ways (unless using real-time PCR, as described below). Typically, a brief purification is performed—as simple as a spin column—to separate the desired product from the unused primers, ultrashort nonspecific products, and remaining free dNTPs. It can then be subjected to gel electrophoresis, stained, and visualized to verify its size, especially in comparison to the positive/negative controls and known molecular weight markers. Alternatively, and perhaps better for most applications, the product can be sequenced or subjected to Southern blot (gel electrophoresis and hybridization with a labeled DNA probe) to confirm its identity. This is necessary in most mutation analysis assays where the sequence is the actual question, and a good quality check for assays where the presence or absence of a product is the desired result; amplification of a nonspecific fragment close to the size of the intended product can lead to spurious results [14].

Advantages/Disadvantages

Advantages

The major advantage of PCR is, of course, its ability to exponentially amplify the target sequence and thus to allow the analysis of very small amounts of input DNA. With the correct choice of primers, PCR is highly sensitive and highly specific. Because of this sensitivity, DNA from a wide range of source materials can be analyzed, and actively growing cells are not required. Current protocols are rapid and easy to perform, taking only hours to perform the reaction and often providing same-day results.

Disadvantages

Similar to FISH (described below), PCR selectively queries a specific sequence of interest and is not designed to be a broad assessment of the genome-like conventional cytogenetics. PCR will also not detect structural rearrangements that leave the target sequence unchanged (although assays can be designed across breakpoints to detect these). PCR can also be subject to variable amplification of particular regions based on sequence and/or structure, which can lead to nonuniform amplification and requires proper validation and controls. Further, the PCR reaction can be inhibited by heparin or melanin if present in the extracted DNA, which may lead to assay failure.

The high sensitivity of PCR can sometimes cause problems; reactions can easily be contaminated by trace amounts of DNA from personnel or other samples. For this reason, molecular diagnostic laboratories must follow strict clean technique with gloves and gowns, filter barriers in pipettors to prevent aerosol contamination, nucleic acid-free consumables, and single-use aliquots of reagents [15]. Laboratories also usually have a unidirectional workflow with separate pre- and post-PCR rooms so that amplified products are never present in the same space where reactions are initially set up. “Wipe tests” should routinely be performed, using swabs from benches and lab surfaces in the various diagnostic tests to ensure that template is not present within the lab.

Applications

PCR is commonly used as a preliminary/preparatory step for other assays, in order to increase the amount of material present, to enrich for specific targets, or to engineer additional sequences onto DNA using specially designed primers. Therefore, many clinical tests and research assays are PCR-based.

PCR can ascertain the presence of a gene or specific sequence. This can be difficult for normal genomic constituents because of the diploid nature of eukaryotic cells; even with a heterozygous loss of 22q, for example, total DNA would be

PCR-positive for the *NF2* gene because of its presence on the second intact copy of 22. PCR works better for novel fusion genes, where primers can be designed to span the breakpoint and will only result in a product if the translocation is present, thus aligning the primers correctly. Identification of fusion genes can be used for tumor diagnosis and often has prognostic significance. In a similar fashion, the clonality of lymphoid proliferations can be assessed by using PCR to look for dominant T-cell receptor or immunoglobulin rearrangement patterns [16, 17].

PCR can also test for point mutations, either by using primers specific for the wild-type or mutant sequence at their 3′ end or by amplifying the region and then sequencing it. Analogously, PCR testing for a panel of single nucleotide polymorphisms (SNPs) can be used for detection of minimal residual disease in leukemia patients, and single tandem repeat analysis can demonstrate engraftment in stem cell transplant recipients [18, 19].

PCR assays can be used, with the addition of chemical modification before the reaction, to determine the methylation status of genomic loci (see below). This can be used to study X-chromosome inactivation patterns to prove (or refute) clonality or to ascertain the inactivation of tumor suppressors in neoplasms, which may have therapeutic or prognostic significance [20, 21]. PCR can also measures alterations in the length of microsatellite sequences in tumors with microsatellite instability [22].

Variations

Reverse Transcriptase PCR (RT-PCR)

Reverse transcriptase PCR is used to amplify RNA sequences of interest by including, as the first step, a reverse transcriptase reaction which creates cDNA from mRNA or another RNA template [23, 24]. cDNA synthesis can be primed by a sequence-specific 3′ oligonucleotide, random hexamer primers, or by using poly(dT) oligonucleotides which will bind to the poly(A) tail of the mRNA and create full-length cDNA. This cDNA then acts at the template in a subsequent conventional or other variant PCR reaction.

These can be performed as entirely separate reactions or as multiple steps within a single tube. Genomic DNA can contaminate the initial RT reaction requiring DNase pretreatment of the input sample, and RNA can contaminate the subsequent PCR, requiring RNase treatment of the synthesized cDNA input. Because of the relative lability of RNA and the extra steps and precautions required when using it as a template, DNA is generally preferred when possible.

RT-PCR is usually used, alone or in combination with other assays, to analyze the sequence or measure the quantity of an RNA target when DNA will not suffice. Such applications include gene expression (mRNA) levels, detection of fusion transcripts or splice variants, or identifying RNA viral sequences such as in microbiology testing [25–27]. Specific mRNA or miRNA profiles can also be used for

detecting minimal residual disease in patients with leukemia or micrometastatic disease or circulating tumor cells in patients with solid tumor patients [28–31].

Real-Time PCR/Quantitative PCR (qPCR)

Real-time PCR has become the most widely used PCR method, because of its accuracy, sensitivity, and ease of (automated) interpretation. The process is quick and measurements (both quantification and detection of target sequences) are complete by the end of the reaction.

There are several issues with using standard (end-point) PCR for quantitation [32]. First and foremost, the amount of product at the end of the reaction correlates poorly with the amount of input DNA. Amplification during a typical PCR begins slowly, becomes exponential during the midphase, and peters out at the end as reagents are exhausted and inhibitory substances are created. Thus, product concentration best mirrors template concentration in the middle of the reaction. Second, it can be difficult to determine absolute DNA concentrations based on control reactions that occur in different tubes due to inherent variability. Third, quantitation can only occur once the reaction cycles are completed and the products are separated by gel electrophoresis. Thus, quantitation by standard PCR is time-consuming and often unreliable.

Real-time PCR circumvents many of these issues by measuring ongoing reaction rates with a PCR reporter that is usually a fluorescent double-stranded DNA binding dye or a fluorescent reporter probe. For instance, real-time PCR assays may utilize probes that exploit the phenomenon of fluorescence resonance energy transfer (FRET) quenching [33]. In this assay format, a pair of oligonucleotide probes labeled with different fluorescent reporter dyes are placed in relatively close proximity. The fluorophore of the probe attached to the 5′ end of the target sequence (donor probe) is excited by an external light source and transfers part of its excitation energy to the adjacent acceptor fluorophore that is attached to the 3′ end of the sequence. The excited acceptor fluorophore emits light at a different wavelength which can then be detected and measured. Fluorescence increases by resonance energy transfer upon hybridization. Interaction of the two fluorophore dyes is distance-dependent and can only occur when both probes are bound to their target, which contributes to increased specificity of the reaction. During the annealing step of a real-time PCR reaction, not only do the forward and reverse primers bind to the ends of the template, but the fluorophore-labeled probe binds in the middle. When hydrolysis probes are utilized, no signal is emitted during the annealing stage of the PCR reaction, because a quencher molecule that is associated with the probe's fluorophore will quench the fluorescence from the fluorophore molecule (Fig. 2.2). But as the DNA polymerase elongates the new strand and reaches the probe, it uses its 5′→3′ exonuclease activity to degrade the probe, thus separating the reporter and the quencher and creating fluorescence. Thus, signal is generated in real time in direct proportion to the number of amplified strands and monitored for each reaction cycle.

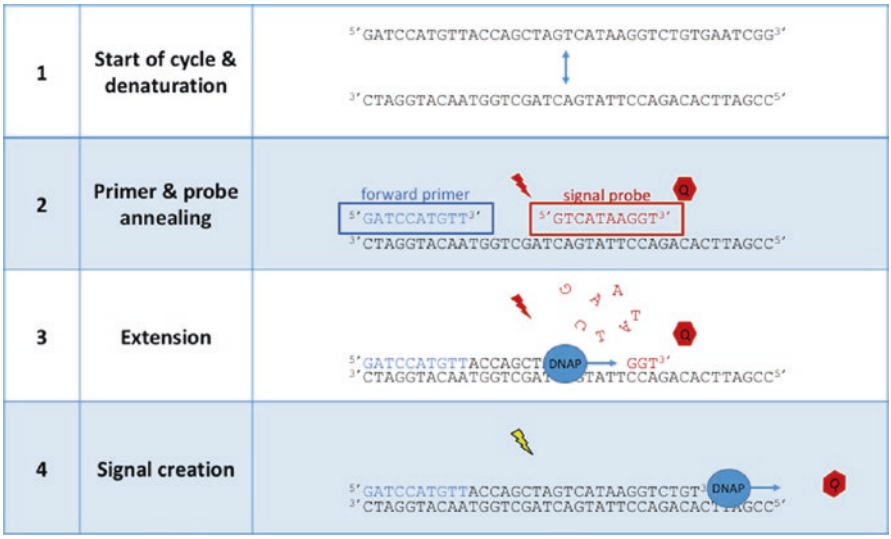


Fig. 2.2 Real-time (quantitative) PCR detection. PCR is performed in the usual fashion, beginning with denaturation of the template and annealing of the primers (1 and 2). However, at this step, a probe complementary to the middle of the target sequence is also annealed to the template (2). There are different dyes and probe formats for real-time PCR. For instance, hydrolysis probes have a fluorescent tag (lightning bolt) at one end and a quencher (hexagon) at the other. When the probe is intact, the quencher prevents any fluorescent signal. As DNA polymerase (DNAP) elongates from the primer, it uses its 5′→3′ exonuclease activity to degrade the probe (3). This separates the fluorophore from the quencher and a signal is produced (4). Thus, signal is directly proportional to the amount of product formed (and, in the log phase of amplification, directly proportional to the amount of template)

Both qualitative and quantitative analyses can be performed by real-time PCR. Qualitative assays use the earliest PCR cycle where signal is detected above background (crossing threshold [Ct] or crossing point [Cp]) as a cutoff for determining the presence or absence of a given target in the reaction. If absolute concentrations are to be calculated, a series of reactions with known amounts of template is used to generate a standard curve to which Ct values of unknown samples are compared. The concentration of the unknown samples is then extrapolated from values from the standard curve as follows: the instrument computer plots the reaction curves (fluorescence vs. cycle number), subtracts out background fluorescence, calculates the threshold cycle (C_t, the earliest cycle where signal is detected) for each reaction, and uses standard formulas to quantify the amount of template present [32].

An additional advantage to real-time PCR is the elimination of post-reaction steps. The measurements are complete when the reaction is, and the sequence specificity of the reporter probe confirms the identity of the product, without the need for gel electrophoresis or sequencing. This is not only a faster alternative to standard PCR but reduces the possibility of contaminating other assays during post-reaction steps (tubes do not have to even be opened post-amplification).

Note that many other public and proprietary variations of this technique exist, most of which differ in the mechanism of their reporter probe methodology, but are beyond the scope of this chapter.

Nested PCR

Nested PCR is a variant of conventional PCR which incorporates a second round of amplification using a second set of internal or "nested" primers. The first PCR reaction is performed in the usual manner, and the amplified product is then used as template in a new reaction. This second reaction has one (semi-nested) or two (nested) different primers which are internal to the original ones; thus, the product from the second reaction is smaller than the first but still designed to contain the region of interest. The two rounds of PCR dramatically increase the degree of amplification, and the two sets of primers increase the specificity. This allows detection of smaller amounts of target while minimizing the amount of nonspecific amplification. Nested PCR is thus commonly used for rare targets or when the source DNA is in low quantity or of low quality.

Multiplex PCR

Multiplex PCR refers to the use of multiple primer sets in a single PCR reaction, thus amplifying several different products of interest at once [34]. The different products may be distinguished by size or by different labels on the primers. The advantages of this technique include increased efficiency and time-savings; it is also useful when only a small amount of template is available. Disadvantages include the often inequitable amplification of the different products due to reactive competition and the unlikelihood that a single set of reaction conditions will be optimal for all the primer sets; thus relative quantitation of the individual products can be tricky. However, with careful planning and optimization, multiplex PCR can be a useful method.

Allele-Specific PCR (ASPCR)

Traditional mutational analysis uses primers to conserved sequences that flank the variable region of interest; the amplified PCR product is sequenced to determine whether a mutation is present. Allele-specific PCR (ASPCR) uses an alternate strategy whereby each primer set is carefully selected to amplify only a single sequence variant [35–37]. This is accomplished by designing the 3′-most end of one primer to be complementary to the variable portion of the sequence. Because it lacks 3′→5′ exonuclease activity, *Taq* polymerase will not extend a DNA strand with a mismatch at the 3′ end, and thus only primers with a perfect end match will be amplified.

Thus, with a set of four reactions (four different forward primers and a single reverse primer) designed to the four possible base choices, the sequence at a given spot can be determined. More commonly, ASPCR panels are designed to identify only the common alleles for a given gene, with appropriate controls run in parallel to ensure that negative reactions represent true sequence mismatches as opposed to technical failures.

ASCPR has found clinical utility in detecting single nucleotide polymorphisms (SNPs) in solid tumor oncogenes and genes associated with constitutional metabolic disorders [38, 39]. Other applications include Rh antigen and HLA genotyping [40].

Digital PCR (dPCR)

Digital PCR involves partitioning the reaction volume into numerous smaller reactions via the use of oil emulsion ("digital droplet"), microwells, or capillary technology [41]. Each of these minireactions contains only a few or no copies of the target sequence. PCR occurs and the signal from each partition is measured individually as positive or negative ("digital" output as opposed to measuring overall signal intensity ["analog"]). Statistical analysis based on the Poisson distribution can then be used to calculate the target prevalence in the overall reaction based on the percentage of positive minireactions. dPCR is well-suited for low copy number targets because the small reaction volume reduces the effects of template competition and nonexponential amplification, the large number of reactions increases reproducibility, and the digital measurements are more accurate than relative fluorescent intensity.

Currently, dPCR is being employed for clinical testing with low-quality DNA samples like formalin-fixed paraffin-embedded tissue and analyzing low-quantity DNA samples such as evaluating mitochondrial DNA heteroplasmy, detecting of circulating tumor cells, and performing the so-called "liquid" biopsies [42–44].

Methylation-Specific PCR

Recently, methylation of CpG islands in genomic DNA has been recognized as a method of inactivating genes not only during normal imprinting but also in a dysregulated fashion in neoplasia [45, 46]. Some malignant tumors have methylation-inactivated tumor suppressor genes, and specific abnormal methylation patterns have been associated with therapeutic response and overall survival.

PCR offers one approach to studying the methylation pattern of tumors [47]. Sodium bisulfite treatment converts unmethylated cytosine to uracil but does not affect methylated cytosine. Thus, this chemical will create different sequences in the plus strand of methylated DNA versus unmethylated DNA and also leave mismatches in the double-stranded DNA (the minus strand is unchanged). By designing

two different forward primers, one against each of the post-treated methylated and unmethylated versions, two PCR reactions can be run to distinguish between them. (Current methylation assays use identical chemistry and rely on changes in restriction enzyme sites caused by the C→U conversion; PCR has the marked advantage of being applicable at any genomic site, not just the ones that are recognized by enzymes.)

Fluorescence In Situ Hybridization (FISH)

Overview

Fluorescence in situ hybridization (FISH) is a molecular technique that allows the identification and localization of specific genetic sequences on chromosomes. FISH uses the hybridization of sequence-specific probes coupled to signaling molecules to identify target genes or regions in their nascent intracellular locations. The simplicity of the procedure, requiring just a few reagents and a fluorescent microscope, and its versatility, probes can be designed against almost any sequence and multiple targets in multiple sources can be tested, have been co-opted for a variety of applications. FISH has found clinical utility in the diagnosis of cancer, constitutional genetic syndromes, and pathogen identification, as well as research utility in molecular genetics and related fields.

Background

The original description of FISH was published in 1982, when biotinylated probes and conjugated antibodies were used to study genes in *Drosophila* chromosomes [48]. Previous attempts had utilized fluorescent signals linked to RNA probes, RNA probes that were detected by anti-DNA-RNA hybrid antibodies, and biotin-conjugated RNA probes [49–52]. One group even used poly(dT)-coated electron-dense spheres to label DNA probes tagged with 3′ poly(dA) sequences [53]. Unfortunately, all these methods had significant shortcomings—probes were short and unstable, and/or autoradiography was insensitive (often requiring days to weeks of exposure), and the signals were difficult to localize within the cells (electron microscopy was often needed).

Langer-Safer et al. used nick translation to label probes and, after hybridizing to the polytene chromosomes, bound a secondary anti-biotin antibody labeled with fluorescein or horseradish peroxidase for detection. This approach represented a significant improvement in ease, resolution, and stability of reagents and would herald the beginning of decades of improved and altered methodology and developing new clinical and research applications for this technique.

Method

The usual FISH protocol requires three components: probe(s), target, and a method to detect their interaction. Different combinations of these define some of the different variants of FISH (discussed below), but the overarching principles remain the same [54].

Probes

FISH probes are usually composed of DNA. Short probes (20–100 nucleotides) can be synthetically constructed using solid-state chemistry to connect bases in the correct order (similar to primer construction for PCR and other applications). Medium-sized probes (100–1000 nucleotides) can be enzymatically assembled using DNA polymerase and a template containing the sequence of interest such as a bacterial plasmid. Longer probes (1000–500,000 nucleotides) can be derived by restriction enzyme digestion and purification of the relevant fragments of a bacterial artificial chromosome (BAC). For common applications, probes and probe sets are commercially available.

The specificity and sensitivity of FISH probes depends on their sequence and length [54]. Longer probes are less likely to have 100% homology to multiple genomic sites, but contain a greater number of short stretches of homology. For example, compared to a 10-base probe (the target of which could randomly occur multiple times in the genome), a 1000-base probe should have a unique genomic match; however, the 1000-base probe itself also contains numerous 10-base sequences with multiple probable matches. Thus, longer probes tend to show more nonspecific binding. Nonetheless, long probes are often necessary to ensure adequate coverage of target sequences and/or to localize enough signal to allow detection. Short probes may be required for higher resolution when mapping a deletion. Most genomic probes utilized for FISH are in the 100–250 kb range, which seems to be the "sweet spot" for maximizing specificity.

The sequence specificity of the probes used in FISH depends on the particular application. For detecting the presence, copy number, or mutational status of a single gene, a locus-specific identifier (LSI) probe is used that is complementary to the unique sequence of that gene or a large portion of it. For counting the number of each chromosome, the so-called centromere enumeration probes (CEPs) are used, which are complementary to repetitive sequences in the center of each chromosome. For the identification of complex rearrangements of portions of a chromosome, whole chromosome painting (WCP) probes are used, with probes sequentially covering each part of a chromosome from one end to the other [55]. (See below for more about these specific applications.)

FISH probes are usually labeled with a fluorophore, an antibody target, or biotin to enable subsequent detection [54]. This can be accomplished in different ways depending on how the probe is manufactured including using modified nucleotides

during synthesis, inserting modified nucleotides post-synthesis, or chemically modifying the ends of the probe post-synthesis. Long probes with signal molecules embedded throughout often suffice for detection, whereas short or end-labeled probes may require signal amplification (see below).

Targets

A wide range of targets are compatible with FISH techniques including primary tumors, cytologic smears or touch preps of solid tumors or leukemias/lymphomas, frozen section slides, formalin-fixed paraffin-embedded (FFPE) tissue sections, and tissue culture cells. For most purposes, cells must be affixed to a solid support. Solid or hematopoietic tumor cells can be touched onto glass slides and fixed in ethanol; cultured cells can be grown directly on glass or plastic coverslips. For frozen section or FFPE slides, thin sections (4–6 μm) typically work best and provide clear, non-overlapping signals. The disadvantage to this is "nuclear truncation"; as nuclei are larger than the section thickness, a given slice will contain an incomplete nucleus and incomplete genomic DNA, and thus an observed "deletion" in a given nucleus could simply be due to the target being out of the plane of section. Thus, numerous nuclei must be scrutinized to ensure a true negative signal. This problem can be circumvented by the use of thick sections (50 μm), but this introduces a new problem of signals being obscured by the thickness of the tissue. Likewise, using disaggregated nuclei also ensures complete signals, but the enzymatic process can damage the nuclei, and the destruction of tissue architecture negates one of the key advantages of FISH.

Whereas RNA targets such as mRNA and miRNA are freely available in the cytoplasm for hybridization, DNA targets are often hidden within heavily packaged chromosomes. This limitation can be overcome by (1) denaturing the DNA and (2) using a target population with a high proportion of cells in interphase or metaphase. For actively growing primary tumor cells in culture or research tissue culture cells, the latter can be achieved by using a mitotic inhibitor such as colchicine that arrests the cells in metaphase; the chromosomes are then optimally spread out for FISH analysis. Alternatively, the high mitotic rate of many tumors may result in enough actively dividing cells at any time such that, at the point of fixation, an adequate number of cells are in interphase or metaphase and easily amenable to FISH.

Recently, circulating tumor cells (CTC) in peripheral blood (a "liquid biopsy") have been used for FISH following their isolation via a commercially available system (CELLSEARCH, Menarini-Silicon Biosystems, San Diego, CA). While this has not yet been performed in the pediatric population, it has shown promise in adult patients with breast or prostate cancer [56]. Briefly, CTC are captured in a cartridge on the basis of a cocktail of tumor-specific antibodies (such as cytokeratin and nonhematopoietic markers), then immobilized on the matrix and used for FISH to detect androgen receptor gene amplification, cancer-specific translocations, and DNA ploidy. This allows the tailoring of therapy in patients with recurrent/progressive cancer without the need for invasive rebiopsy.

Procedure

Although there are many different protocols for performing FISH, the steps are similar and most assays are completed within 1–2 days (Fig. 2.3). Target cells are prepared according to their source: touch preps, cytologic smears, or tissue culture slides are fixed with alcohol or formalin and permeabilized with alcohol or detergent; FFPE sections are deparaffinized and rehydrated in buffer. Probe and target DNA are then denatured at high temperature in a suitable buffer, creating single-stranded nucleic acid suitable for binding. Hybridization is then allowed to occur; the optimal temperature will depend on the probe, and the length of time can range from 4 h to overnight (again depending on the length and sequence of the probe). A preincubation with short unlabeled random DNA probes can be used to block nonspecific binding of the actual probes, especially to long repetitive sequences. After hybridization, several washes are performed at increasingly stringent conditions (higher salt concentrations and warmer temperatures) to remove unbound and nonspecifically bound probes. Nuclei are counterstained, often with 4′6′-diamidino-2-phenylindole (DAPI), and the signal is then visualized.

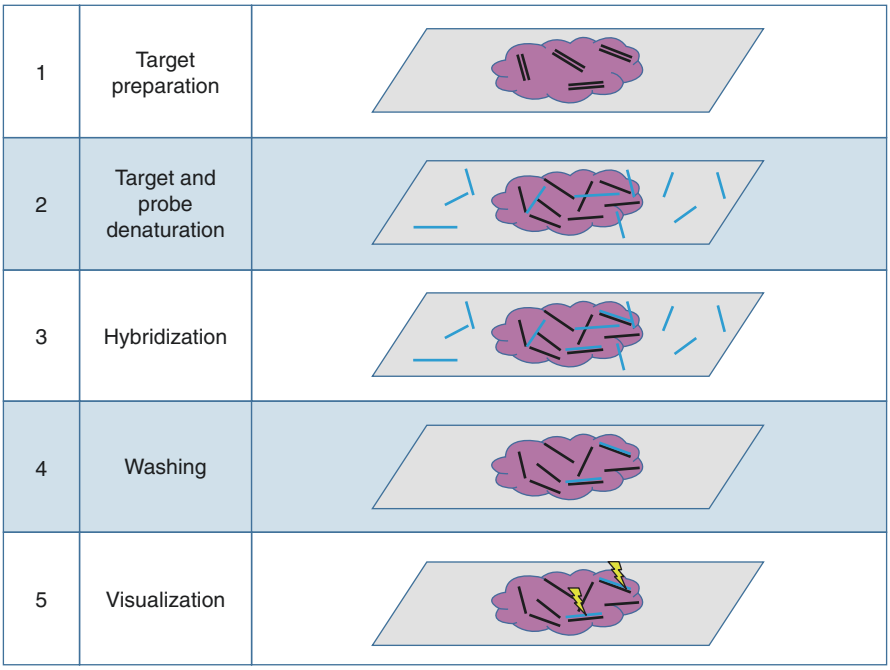


Fig. 2.3 Overview of standard FISH technique. The target, in this example a section of formalin-fixed paraffin-embedded tumor tissue, is prepared by deparaffinization and retrieval with enzyme and/or boiling buffer to remove DNA-protein complexes (1). Target and labeled probe DNA are then denatured by heat treatment to create single strands (2). They are allowed to hybridize from 4 h to overnight depending on the assay (3). Unbound and nonspecifically bound probes are then washed away (4). Finally, the signal is visualized using a fluorescent microscope (5)

Signals and Detection

Visualization of hybridized fluorescent probes can usually occur directly with a fluorescence microscope, because of the high signal strength, labeling density, and signal-noise ratio of these systems. Multiple fluorescent probes can be employed in a single assay thus allowing the simultaneous identification of several probes (such as in fusion detection or whole chromosome painting). Chromogenic in situ hybridization (CISH) uses biotin or digoxigenin and secondary detection reagents such as streptavidin-conjugated horseradish peroxidase (HRP) to create brown staining which can be seen with a traditional bright-field microscope, analogous to immunohistochemistry. CISH is cheaper and procedurally easier to detect (any lab performing immunohistochemistry should already be well-equipped for CISH), and the resulting signals are near-permanent, unlike the eventually fading fluorescence of traditional FISH. (Most labs capture digital images of FISH assays for permanent record-keeping as signals will fade within months even under ideal storage conditions.) However, CISH signals are often less intense, less discrete, and have a lower signal-noise ratio. Double-staining CISH uses antibodies against fluorescent dyes and conjugated to HRP or other enzymes in order to convert a FISH system into CISH [57].

More recently, alternate signaling platforms have been developed to increase sensitivity and ease of use. Van Gijlswijk et al. have reported on the use of tyramide as an HRP substrate (tyramide signal amplification, or TSA; [58]). In this method, a tyramide-conjugated biotin- or fluorescent-based marker is used. At the site of HRP-probe binding, the HRP converts the tyramide into a highly reactive radical which forms a covalent linkage to the target molecule at or near that site. The biotin or fluorescence are then detected in the usual way. TSA is an example of catalyzed reporter deposition (CARD), which serves to amplify the signal thereby lowering detection limits and allowing the use of shorter or lower concentrations of probes.

Advantages and Disadvantages

Advantages

The major advantage to FISH, as opposed to traditional in vitro detection versions of PCR and Southern blots, is the ability to localize the gene of interest to a particular chromosome, chromosomal region, or proximity to another gene. FISH can also be performed on fixed tissue sections and cells unlike conventional cytogenetic analysis which requires viable, actively dividing cells. FISH allows the evaluation of specific subpopulations of cells in the setting of their nascent morphologic architecture without the need for microdissection and without the concerns of normal cells outgrowing the cells of interest in culture. FISH has higher resolution than karyotyping and banding, and is superior for detection of small deletions.

Disadvantages

Traditional FISH methods are not without their limitations. FISH requires the selection of a specific target sequence that is queried, whereas conventional cytogenetics broadly assess the whole genome. FISH has relatively low analytical sensitivity compared to target amplification methods like PCR but is often more specific depending on the application.

FISH is susceptible to many of the technical issues seen in other nucleic acid assays—false positives and negatives, the inability to use decalcified tissue, and staining artifacts. In order to confirm that an assay has worked properly (especially when identifying a deletion) and to prevent partial hybridization and nuclear truncation errors, control probes and/or control cells must be run each time. Copy number or ploidy studies also require a reference probe to normalize the results of each examined cell. One lingering issue, unrelated to the method, is the lack of generally accepted definitions and criteria for amplifications and deletions; standardization would facilitate clinical and basic science research applications of FISH.

Applications

Chromosomal Aneusomies

FISH can readily be utilized to determine the copy number of each chromosome in a cell's genome. While this is most often used for studying constitutional genetic syndromes, it has found several applications in pediatric cancer. For example, hyperdiploid status is a favorable prognosis in neuroblastoma, and specific trisomies confer favorable or unfavorable prognosis in pediatric leukemias. The easiest approach is to use CEPs for each chromosome, with 3–5 chromosomes examined in each method; copy numbers can thus be assessed. However, there is significant cross-reactivity between the CEPs of some chromosomes, which can hinder this approach. Thus, some laboratories will use multiple LSIs for specific genes on each chromosome or will use WCPs to target all/most of the chromosome instead. Another option is subtelomeric probes, which while still consisting of repetitive sequences may show less cross-hybridization between different chromosomes [59].

Segmental Deletions

Segmental chromosomal aberrations convey useful diagnostic and prognostic information in pediatric tumors. Pediatric glioma patients with codeletion of 1p and 19q have better response to therapy and a more hopeful prognosis [60]. Also, loss of 22q is a common finding in malignant rhabdoid tumors and correlates with loss of the *INI1/hSNF5* gene. To assess the integrity of a specific chromosome, multiple LSIs can be employed as a surrogate for chromosomal segmental deletion (e.g., both *NF2*

and *BCR* are located on 22q; absence of one or both of these genes suggests 22q deletion). Alternatively, WCP probes can be used to more precisely map the location and extent of a deletion. In either case, CEPs or LSIs on the opposite chromosomal arm should be used as controls to verify that the region's absence is due to a partial chromosomal deletion and not monosomy or another aneusomy.

Single Gene Amplification

The quantitative nature of FISH is quite useful for detecting gene amplification. One of the most important applications of this technique in pediatric cancer is the assessment of *MYCN* amplification in neuroblastoma, which is associated with high-risk disease. Gene amplifications typically occur in one of two ways: (1) homogeneously staining regions (HSR), where multiple copies of a single gene lie in a row on the chromosome, and (2) double minutes or small extrachromosomal DNA fragments containing the gene and numbering in the tens to hundreds (the usual case in neuroblastoma). To detect either mode of amplification, an LSI probe specific for the gene of interest is used in conjunction with a CEP or other control probe (a specific gene on the same chromosome as the LSI); this is crucial to distinguish true amplification from polysomy and to accurately quantify the degree of amplification..

Translocations

Another common use for FISH is in detecting chromosomal translocations, especially those associated with particular tumors. Gene rearrangements such as *EWS-FLI1* in Ewing sarcoma, *PAX3-FOXO1* in alveolar rhabdomyosarcoma, and *MYCC-IGH* in Burkitt lymphoma serve as useful diagnostic adjuncts if not pathognomonic findings. In this application of FISH, the key information regards the relative location of the probe(s)/signal(s), not the quantification. Two different approaches can be used: identifying the fusion gene or identifying the separation of one of the involved genes.

In "fusion" FISH, two probes with different signals are used, one specific for each member of the translocated gene. In a normal cell, both copies of each gene are present in their usual and separate locations, so four distinct signals are seen. When a translocation is present, two of the genes come together, and thus two of the signals co-localize; the two on the uninvolved genes remain separate. Depending on the particular probes and signals, the juxtaposition may form a new signal color (with standard fluorescent probes, red and green together become yellow). This technique requires both partner genes to be known, and a rearrangement involving only one of the candidate genes will not be detected. For example, fusion FISH using *EWS* and *FLI1* probes will identify the most common *EWS-FLI1* rearrangement of Ewing sarcoma but will not identify *EWS-ERG* or the other less common fusion genes. One important caveat is that just by chance, the orientation of nuclei and chromosomes in the tissue preparation will on occasion juxtapose the two sig-

nals even when they are on separate chromosomes. For this reason, it is important to examine numerous nuclei and to only report a gene fusion when a significant percentage of cells have the same finding.

The second technique, "breakapart" FISH, utilizes two probes at opposite sides of the breakpoint within a single gene. In the normal state, these signals from these two probes co-localize and/or form a secondary color. If the gene is involved in a rearrangement at a breakpoint between the probes, the probes separate and two separated signals will be observed (in addition to the other, nonrearranged gene signal). This technique is broader than fusion FISH and allows detection of a group of translocations involving a single gene (such as the Ewing sarcoma family of tumors involving different *EWS* translocations—in this case, probes at the 5′ and 3′ ends of the EWS gene would detect both the *EWS-FLI1* and *EWS-ERG* rearrangements), but additional workup through further FISH or PCR may be required to identify the partner gene. The target location of probes can be chosen to be specific for a single breakpoint in the gene or can straddle a cluster of multiple breakpoints depending on the desired information. Because nuclear or chromosomal positioning cannot mimic a rearrangement, breakapart FISH has a lower false-positive rate than fusion FISH, and the threshold for calling a positive is lower. One group advocates a threshold of 30% fused signals or 15% broken-apart signals before calling the respective assays positive [61].

Both these methods can be modified in several ways. Most commonly, additional control or "marker" probes will be added. For example, using a third probe specific for the centromeric region of the presumed destination chromosome may localize the new fusion gene and further support its identity; this helps lower the possibility of a false positive from overlying chromosomes in fusion FISH or may help suggest a fusion partner in breakapart FISH. Nonetheless, in the presence of complex rearrangements, FISH with multiple probes can be difficult to interpret and additional studies may be required.

Other Clinically Relevant FISH Variants

CODFISH (concomitant oncoprotein detection with FISH) combines standard FISH for gene amplification with immunohistochemistry for expression of the protein product of the same gene. It utilizes a three-color system for the target gene, the control CEP, and a unique immunohistochemistry chromogen that can be detected by both bright-field and fluorescence microscopy. It was originally described for evaluating Her-2/*neu* in breast carcinoma [62].

FICTION (fluorescence immunophenotyping and interphase cytogenetics as a tool for investigation of neoplasms) uses immunophenotyping with fluorescent-conjugated antibodies against cell surface proteins in conjunction with FISH for chromosomal aneusomies. This has mostly been used for leukemia/lymphoma and myelodysplastic syndromes as it obviates the need for flow cytometry prior to FISH [63, 64].

Flow FISH is a variant of Q-FISH (below) that first performs probe hybridization on leukocytes in suspension and then sorts the cells via a flow cytometer to measure telomere (or other repetitive sequence) length in subsets of leukocytes.

Q-FISH (quantitative FISH) uses synthetic probes with a peptide-nucleic acid backbone (PNA oligonucleotides). PNA oligonucleotides lack the negatively charged phosphate groups of normal DNA and hybridize under more stringent conditions than normal probes; this increased specificity makes the assay more sensitive and more quantitative. Q-FISH is primarily used to measure telomere lengths.

SKY (spectral karyotyping) is a combination of whole chromosome painting FISH and conventional karyotyping. A large set of probes is generated that covers each chromosome, with a unique fluorophore for each chromosome. When the probes are hybridized to a metaphase spread, each chromosome is painted a different color. This facilitates identification of each chromosome and highlights any translocations as nonmatching colors. SKY is a type of *multiplex FISH (M-FISH)*.

smRNA-FISH (single-molecule FISH) uses multiple short-labeled oligonucleotides as probes to measure the localization and expression of mRNA within cells.

TMA FISH (tissue microarray FISH) is standard FISH performed on tissue microarray slides instead of whole sections. It can correlate immunohistochemistry and FISH findings on serial sections but requires multiple cores from each tumor to ensure faithful representation. TMA FISH is useful for screening or for validating probes and is quite amenable to computer-based image analysis.

References

1. Kleppe K, Ohtsuka E, Kleppe R, Molineux I, Khorana HG. Studies on polynucleotides. XCVI. Repair replications of short synthetic DNAs as catalyzed by DNA polymerases. J Mol Biol. 1971;56:341–61.
2. Mullis KB, Faloona FA. Specific synthesis of DNA *in vitro* via a polymerase-catalyzed chain reaction. Methods Enzymol. 1987;155:335–50.
3. Saiki RK, et al. Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. Science. 1988;239:487–91.
4. Impraim CC, Saiki RK, Erlich HA, Teplitz RL. Analysis of DNA extracted from formalin-fixed, paraffin-embedded tissues by enzymatic amplification and hybridization with sequence-specific oligonucleotides. Biochem Biophys Res Commun. 1987;142(3):710–6.
5. Ren ZP, Sällström J, Sundström C, Nistér M, Olsson Y. Recovering DNA and optimizing PCR conditions from microdissected formalin-fixed and paraffin-embedded materials. Pathobiology. 2000;68(4–5):215–7.
6. Greer CE, Lund JK, Manos MM. PCR amplification from paraffin-embedded tissues: recommendations on fixatives for long-term storage and prospective studies. PCR Methods Appl. 1991;1(1):46–50.
7. Karlsen F, Kalantari M, Chitemerere M, Johansson B, Hagmar B. Modifications of human and viral deoxyribonucleic acid by formaldehyde fixation. Lab Investig. 1994;71(4):604–11.
8. Wickham CL, Sarsfield P, Joyner MV, Jones DB, Ellard S, Wilkins B. Formic acid decalcification of bone marrow trephines degrades DNA: alternative use of EDTA allows the amplification and sequencing of relatively long PCR products. Mol Pathol. 2000;53(6):336. Erratum in: Mol Pathol 2001 Apr;54(2):120.

9. Farkas DH, Drevon AM, Kiechle FL, DiCarlo RG, Heath EM, Crisan D. Specimen stability for DNA-based diagnostic testing. Diagn Mol Pathol. 1996;5(4):227–35. Erratum in: Diagn Mol Pathol 1997 Jun;6(3):178.
10. Visvikis S, Schlenck A, Maurice M. DNA extraction and stability for epidemiological studies. Clin Chem Lab Med. 1998;36(8):551–5.
11. Hung JH, Weng Z. Designing polymerase chain reaction primers using Primer3Plus. Cold Spring Harb Protoc. 2016;2016(9):pdb.prot093096.
12. Hommelsheim CM, Frantzeskakis L, Huang M, Ülker B. PCR amplification of repetitive DNA: a limitation to genome editing technologies and many other applications. Sci Rep. 2014;4:5052.
13. Terpe K. Overview of thermostable DNA polymerases for classical PCR applications: from molecular and biochemical fundamentals to commercial systems. Appl Microbiol Biotechnol. 2013;97(24):10243–54.
14. Pfeifer JD. Polymerase chain reaction. In: Pfeifer JD, editor. Molecular genetic testing in surgical pathology. Philadelphia: Lippincott Williams & Wilkins; 2006.
15. Kwok S, Higuchi R. Avoiding false positives with PCR. Nature. 1989;339(6221):237–8. Erratum in: Nature 1989 Jun 8;339(6224):490.
16. Langerak AW, Groenen PJ, Brüggemann M, Beldjord K, Bellan C, Bonello L, et al. EuroClonality/BIOMED-2 guidelines for interpretation and reporting of Ig/TCR clonality testing in suspected lymphoproliferations. Leukemia. 2012;26(10):2159–71.
17. Rezuke WN, Abernathy EC, Tsongalis GJ. Molecular diagnosis of B- and T-cell lymphomas: fundamental principles and clinical applications. Clin Chem. 1997;43(10):1814–23.
18. Matsuda K, Sugano M, Honda T. PCR for monitoring of minimal residual disease in hematologic malignancy. Clin Chim Acta. 2012;413(1–2):74–80.
19. Van Deerlin VM, Leonard DG. Bone marrow engraftment analysis after allogeneic bone marrow transplantation. Clin Lab Med. 2000;20(1):197–225.
20. Diaz-Cano SJ, Blanes A, Wolfe HJ. PCR techniques for clonality assays. Diagn Mol Pathol. 2001;10(1):24–33.
21. Kristensen LS, Hansen LL. PCR-based methods for detecting single-locus DNA methylation biomarkers in cancer diagnostics, prognostics, and response to treatment. Clin Chem. 2009;55(8):1471–83.
22. Berg KD, Glaser CL, Thompson RE, Hamilton SR, Griffin CA, Eshleman JR. Detection of microsatellite instability by fluorescence multiplex polymerase chain reaction. J Mol Diagn. 2000;2(1):20–8.
23. Jensen EC. Real-time reverse transcription polymerase chain reaction to measure mRNA: use, limitations, and presentation of results. Anat Rec (Hoboken). 2012;295(1):1–3.
24. Mayer G, Müller J, Lünse CE. RNA diagnostics: real-time RT-PCR strategies and promising novel target RNAs. Wiley Interdiscip Rev RNA. 2011;2(1):32–41.
25. Bridge JA. Reverse transcription-polymerase chain reaction molecular testing of cytology specimens: pre-analytic and analytic factors. Cancer. 2017;125(1):11–9.
26. Merckx J, Wali R, Schiller I, Caya C, Gore GC, Chartrand C, et al. Diagnostic accuracy of novel and traditional rapid tests for influenza infection compared with reverse transcriptase polymerase chain reaction: a systematic review and meta-analysis. Ann Intern Med. 2017;167(6):394–409.
27. Wilkin F, Gagné N, Paquette J, Oligny LL, Deal C. Pediatric adrenocortical tumors: molecular events leading to insulin-like growth factor II gene overexpression. J Clin Endocrinol Metab. 2000;85(5):2048–56.
28. Beiske K, Ambros PF, Burchill SA, Cheung IY, Swerts K. Detecting minimal residual disease in neuroblastoma patients-the present state of the art. Cancer Lett. 2005;228(1–2):229–40.
29. Lianidou ES, Markou A. Molecular assays for the detection and characterization of CTCs. Recent Results Cancer Res. 2012;195:111–23.
30. Schrappe M. Detection and management of minimal residual disease in acute lymphoblastic leukemia. Hematology Am Soc Hematol Educ Program. 2014;2014(1):244–9.

31. Vo KT, Edwards JV, Epling CL, Sinclair E, Hawkins DS, Grier HE, et al. Impact of two measures of micrometastatic disease on clinical outcomes in patients with newly diagnosed ewing sarcoma: a report from the children's oncology group. Clin Cancer Res. 2016;22(14):3643–50.
32. Arya M, Shergill IS, Williamson M, Gommersall L, Arya N, Patel HR. Basic principles of real-time quantitative PCR. Expert Rev Mol Diagn. 2005;5(2):209–19.
33. Faner R, Casamitjana N, Colobran R, Ribera A, Pujol-Borrell R, Palou E, et al. HLA-B27 genotyping by fluorescent resonance emission transfer (FRET) probes in real-time PCR. Hum Immunol. 2004;65(8):826–38.
34. Edwards MC, Gibbs RA. Multiplex PCR: advantages, development, and applications. PCR Methods Appl. 1994;3(4):S65–75.
35. Bottema CD, Sommer SS. PCR amplification of specific alleles: rapid detection of known mutations and polymorphisms. Mutat Res. 1993;288(1):93–102.
36. Matsuda K. PCR-based detection methods for single-nucleotide polymorphism or mutation: real-time PCR and its substantial contribution toward technological refinement. Adv Clin Chem. 2017;80:45–72.
37. Ugozzoli L, Wallace RB. Allele-specific polymerase chain reaction. Methods. 1991;2(1):42–8.
38. Quiros RM, Ding HG, Gattuso P, Prinz RA, Xu X. Evidence that one subset of anaplastic thyroid carcinomas are derived from papillary carcinomas due to BRAF and p53 mutations. Cancer. 2005;103(11):2261–8.
39. Wilson RC, Wei JQ, Cheng KC, Mercado AB, New MI. Rapid deoxyribonucleic acid analysis by allele-specific polymerase chain reaction for detection of mutations in the steroid 21-hydroxylase gene. J Clin Endocrinol Metab. 1995;80(5):1635–40.
40. Le Van Kim C, Colin Y, Brossard Y, Cartron JP. Rh haemolytic disease of the newborn and Rh genotyping by RFLP--and allele-specific--PCR. Transfus Clin Biol. 1995;2(4):317–24.
41. Vogelstein B, Kinzler KW. Digital PCR. Proc Natl Acad Sci U S A. 1999;96(16):9236–41.
42. Huang A, Zhang X, Zhou SL, Cao Y, Huang XW, Fan J, Yang XR, Zhou J. Detecting circulating tumor DNA in hepatocellular carcinoma patients using droplet digital PCR is feasible and reflects intratumoral heterogeneity. J Cancer. 2016;7(13):1907–14.
43. Olmedillas-López S, García-Arranz M, García-Olmo D. Current and emerging applications of droplet digital PCR in oncology. Mol Diagn Ther. 2017;21(5):493–510.
44. Sofronova JK, Ilinsky YY, Orishchenko KE, Chupakhin EG, Lunev EA, Mazunin IO. Detection of mutations in mitochondrial DNA by droplet digital PCR. Biochemistry (Mosc). 2016;81(10):1031–7.
45. Baylin SB, Makos M, Wu JJ, Yen RW, de Bustros A, Vertino P, et al. Abnormal patterns of DNA methylation in human neoplasia: potential consequences for tumor progression. Cancer Cells. 1991;3(10):383–90.
46. Goyama S, Kitamura T. Epigenetics in normal and malignant hematopoiesis: an overview and update 2017. Cancer Sci. 2017;108(4):553–62.
47. Herman JG, Graff JR, Myöhänen S, Nelkin BD, Baylin SB. Methylation-specific PCR: a novel PCR assay for methylation status of CpG islands. Proc Natl Acad Sci U S A. 1996;93(18):9821–6.
48. Langer-Safer PR, Levine M, Ward DC. Immunological method for mapping genes on Drosophila polytene chromosomes. Proc Natl Acad Sci U S A. 1982;79(14):4381–5.
49. Bauman JG, Wiegant J, Borst P, van Duijn P. A new method for fluorescence microscopical localization of specific DNA sequences by in situ hybridization of fluorochrome labelled RNA. Exp Cell Res. 1980;128(2):485–90.
50. Cheung SW, Tishler PV, Atkins L, Sengupta SK, Modest EJ, Forget BG. Gene mapping by fluorescent in situ hybridization. Cell Biol Int Rep. 1977;1(3):255–62.
51. Manning JE, Hershey ND, Broker TR, Pellegrini M, Mitchell HK, Davidson N. A new method of in situ hybridization. Chromosoma. 1975;53(2):107–17.
52. Rudkin GT, Stollar BD. High resolution detection of DNA-RNA hybrids in situ by indirect immunofluorescence. Nature. 1977;265(5593):472–3.

53. Wu M, Davidson N. Transmission electron microscopic method for gene mapping on polytene chromosomes by in situ hybridization. Proc Natl Acad Sci U S A. 1981;78(11):7059–63.
54. Bayani J, Squire JA. Fluorescence in situ hybridization (FISH). Curr Protoc Cell Biol. 2004;Chapter 22:Unit 22.4.
55. Tucker JD. Reflections on the development and application of FISH whole chromosome painting. Mutat Res Rev Mutat Res. 2015;763:2–14.
56. Swennenhuis JF, Terstappen L. Sample preparation methods following cellsearch approach compatible of single-cell whole-genome amplification: an overview. Methods Mol Biol. 2015;1347:57–67.
57. van Rijk A, Svenstroup-Poulsen T, Jones M, Cabeçadas J, Cigudosa JC, Leoncini L, et al. Double-staining chromogenic in situ hybridization as a useful alternative to split-signal fluorescence in situ hybridization in lymphoma diagnostics. Haematologica. 2010;95(2):247–52.
58. van Gijlswijk RP, van de Corput MP, Bezrookove V, Wiegant J, Tanke HJ, Raap AK. Synthesis and purification of horseradish peroxidase-labeled oligonucleotides for tyramide-based fluorescence in situ hybridization. Histochem Cell Biol. 2000;113(3):175–80.
59. Knight SJ, Flint J. Perfect endings: a review of subtelomeric probes and their use in clinical diagnosis. J Med Genet. 2000;37(6):401–9.
60. Ducray F, Idbaih A, Wang X-W, Cheneau C, Labussiere M, Sanson M. Predictive and prognostic factors for gliomas. Expert Rev Anticancer Ther. 2011;11(5):781–9.
61. Bridge RS, Rajaram V, Dehner LP, Pfeifer JD, Perry A. Molecular diagnosis of Ewing sarcoma/primitive neuroectodermal tumor in routinely processed tissue: a comparison of two FISH strategies and RT-PCR in malignant round cell tumors. Mod Pathol. 2006;19(1):1–8.
62. Tubbs RR, Pettay J, Roche P, et al. Concomitant oncoprotein detection with fluorescence *in situ* hybridization (CODFISH): a fluorescence-based assay enabling simultaneously visualization of gene amplification and encoded protein expression. J Mol Diagn. 2000;2:78–83.
63. Weber-Matthiesen K, Deerberg J, Poetsch M, Grote W, Schlegelberger B. Numerical chromosome aberrations are present within the CD30+ Hodgkin and Reed-Sternberg cells in 100% of analyzed cases of Hodgkin's disease. Blood. 1995;86(4):1464–8.
64. Nolte M, Werner M, Ewig M, et al. Fluorescence in situ hybridization (FISH) is a reliable diagnostic tool for detection of the 9;22 translocation. Leuk Lymphoma. 1996;22(3–4):287–94.

Chapter 3
Molecular Methods in Oncology: Genomic Analysis

Jason A. Jarzembowski

Next-Generation Sequencing

Overview

The improvements in sequencing technology over the past 40 years since the first genomic DNA sequence was published have been astounding; the human genome that took 13 years and millions of dollars to sequence when the project started in 1990 today takes less than a week and costs only a few thousand dollars. Advances in semiconductors, microfluidics, optics, and computing have all contributed to this success, along with the hard work and ingenuity of many scientists and engineers. The question is no longer "can we sequence a patient's constitutional or tumor genome?" but "should we?" and "what will we do with that information?"

Background

Fred Sanger and colleagues described their chain-termination method of DNA sequencing ("Sanger" or "first-generation" sequencing) in 1977 and used it to re-sequence the 5375-base genome of the bacteriophage ϕX174, which had already been determined by other methods [1, 2]. This method utilized dideoxynucleotides (ddNTPs) which could be incorporated into DNA by a polymerase but not further extended because ddNTPs lack the 3′ hydroxyl group required for extension of DNA chains. Thus, a reaction containing a single-stranded DNA template, a primer complementary to one end, DNA polymerase, dNTPs (including dATP), and ddATP

J. A. Jarzembowski, MD, PhD
Division of Pediatric Pathology, Medical College of Wisconsin, Milwaukee, WI, USA
e-mail: jjarzemb@mcw.edu

L. V. Furtado, A. N. Husain (eds.), *Precision Molecular Pathology of Neoplastic Pediatric Diseases*, Molecular Pathology Library, https://doi.org/10.1007/978-3-319-89626-7_3

would result in DNA fragments complementary to the template, all with identical 5′ ends, and with assorted lengths, all terminating at a 3′-ddATP somewhere where the template sequence contained a "T." Thus, if four parallel reactions are run, each with a different ddNTP, each possible fragment length should be present within one of the tubes. By incorporating a radioactive or fluorescent label into the primer or one of the dNTPs and resolving the reactions by acrylamide gel electrophoresis, it is possible to identify the terminal nucleotide for each position in the fragment, thereby uncovering the DNA sequence.

The Sanger method quickly became the most widely used sequencing technique, and it was the approach used at the beginning of the Human Genome Project in 1990 [3]. Because Sanger sequencing is usually limited to reads of <1000 bases, the 3.3 billion base pair genome was first broken into 150 kb segments. Each of these was sequenced using a "shotgun" approach—the 150 kb segments were then randomly broken into smaller pieces and sequenced. The fragmentation and sequencing process was then repeated multiple times to create numerous chunks of sequence, which were then reassembled by a computer based on overlap to reconstitute the 150 kb segment. All the 150 kb segments were then pieced together to form the genome sequence; the Human Genome Project finished in 2003, 2 years ahead of schedule.

Along the way, Sanger sequencing became easier and faster. In the mid-1980s, fluorescent technology allowed the usage of different color signals on each of the ddNTPs, which meant that all four nucleotides could be included in a single reaction tube [4, 5]. Fluorescence also eliminated the need for radioactive reagents, reducing the hazard, cost, and hassle of using this material. Capillary electrophoresis soon replaced gel electrophoresis in automated sequencing machines; almost 100 samples could be run in each batch in just a few hours. Automated sequencers became the workhorse for the Human Genome Project and for labs everywhere.

Methods

Next-Generation Sequencing (NGS)

"Next-generation" sequencing is the name given to the successors to Sanger sequencing and is sometimes and more properly perhaps called "second generation" or simply "high throughput" [6–8]. The length limit on Sanger sequencing (around 1 kb) occurs because it becomes increasingly difficult to distinguish between two large DNA molecules that only differ by one base by gel or capillary electrophoresis, or even other technologies. Thus, NGS methods largely accept that limit and work with a brute force approach to accomplish the task: rather than sequence one long fragment, they break the DNA into thousands or millions of small fragments, sequence those in a massively parallel fashion, and then piece the sequences together with sheer computational will. As described below, most of these approaches employ "sequencing by synthesis," meaning (like Sanger) they elongate strands with DNA polymerase and track which bases get added; the single exception is

Table 3.1 Comparison of next-generation sequencing platforms

Platform	Unique chemistry	Read lengths	Run time
Solexa/Illumina	Bridge amplification; reversible terminators	Up to 150 bases	4–48 h
Ion semiconductor	ISFET detection	200–400 bases	2–4 h
SMRT	Single-molecule detection with ZMV	10,000 bases	1–96 h
SOLiD	DNA ligase, octamer reading of alternating sites	50–60 bases	6 days

sequencing by oligonucleotide ligation and detection (SOLiD), which uses DNA ligase instead. The commonalities of the NGS techniques listed below outweigh their differences, and while each has its strengths and weaknesses, all can be adapted to work for most clinical and research purposes (Table 3.1).

Solexa/Illumina

This methodology begins with genomic DNA fragmented into about 200 bp pieces [9, 10]. Adapters are ligated onto both ends; these contain terminal sequences, primer binding sites, and index sites (each fragment has two different adapters). The adapter-ligated DNA molecules are put into a flow cell, where the terminal sequences bind to immobilized oligonucleotides on the reaction surface. Localized PCR is performed to generate DNA "clusters" of the forward and reverse strands of the original fragment (since either end of the fragment can bind to the surface, and the hybridized portion serves as a PCR primer). Replicated molecules denature and bend to attach their other end to the surface, and replication begins anew ("bridge amplification"). Sequences are read by incorporation of modified nucleotides that have a conjugated dye and a reversible terminator [11]. The signals are read to identify the single base added in each cluster; then, the dye and terminator are chemically removed from the product, and the cycle is repeated. After the forward strands have been read, the reverse strands are synthesized using primers and are then sequenced in the same way. The analysis of sense and antisense strands serves as an internal control as the two sequences can be compared against each other. The index sites allow multiple samples to be run together but analyzed separately. Sequence reads are typically 50–150 bases long, requiring alignment to reconstruct the genomic sequence. The large number of molecules analyzed provides excellent coverage, and this is one of the more popular systems.

Ion Semiconductor Sequencing

Fragmented DNA (<500 bp) is ligated to adaptors and mixed with beads in a dilute enough concentration to bind one DNA molecule per bead [10, 12]. Emulsion PCR is performed to cover each bead with copies of a single DNA fragment. (Emulsion PCR takes places in an oil/water suspension, such that individual reactions occur in

separate miniscule aqueous bubbles without mixing until the suspension is disrupted [13].) The beads are dispensed into individual picoliter-sized wells (1–100 million/chip), each of which has an ion-sensitive field-effect transistor (ISFET, functioning as a solid-state pH monitor [14]). The wells are filled with buffer containing DNA polymerase and a single type of dNTP. If the open position in the sequence in a given well calls for that type of dNTP, a base is added; if several identical bases occur in the sequence in a row, multiple dNTPs are added in this step. The chemistry of base addition releases pyrophosphate and a hydrogen ion; the number of hydrogen ions released in a well is proportionate to the number of dNTPs added, and these are detected as electric pulses by the ISFET. If the open position in the sequence in a given well does not call for that particular dNTP, nothing happens. After data recording, the wells are drained and washed, and the cycle is repeated with new dNTPs, rotating through all four dNTPs. Because of the variable appearance of each of the bases in any given DNA fragment, the resulting sequences will be of variable length.

This system is relatively inexpensive because it avoids the use of specially modified nucleotides and the need for optical measuring instruments (supplanted by the ISFET). The disadvantages are small sequence runs (about 200–400 bp) that require some bioinformatics work to reassemble and difficulty in accurately reading sequence repeats (homopolymers) of more than five bases. Thus, ion semiconductor sequencing is good for sequences which will be aligned to known reference standards, which speeds analysis and decreases possible errors.

Single-Molecule Real-Time (SMRT) Sequencing

This platform has 100,000–1,000,000 zero-mode waveguides (ZMV) per chip, each of which is a zeptoliter volume area with an immobilized DNA polymerase/DNA template complex [15, 16]. Fluorophore-conjugated nucleotides (each dNTP is linked to a different color dye via the phosphate moiety) are added and elongation begins. As a nucleotide is added, the fluorophore is released and recorded; because of the miniscule size of the detection volume, the dye quickly diffuses out of the optical path before the next nucleotide is added.

The average read length of SMRT is on the order of 10,000 bases, well in excess of standard Sanger sequencing and most NGS variants; this makes reassembly and alignment of the data easier. Thus, SMRT is ideal for de novo genome sequencing, detecting isoforms, and sequencing repetitive elements. However, the technology and specialized reagents are more expensive and have a higher error rate per base than other platforms.

Sequencing by Oligonucleotide Ligation and Detection (SOLiD)

SOLiD begins with a fragmented genomic DNA library that is capped with common ends (via DNA ligase) and immobilized to beads or a slide, depending on the exact platform [10, 17]. Local amplification takes place via emulsion or solid-state

PCR to form clusters of identical template. Next, an oligonucleotide complementary to the common primer is bound. Then, fluorescently labeled octamer probes are bound next to the first. These are chemically degenerate such that alignment only needs to occur with the central one to three bases (depending on the version); each probe has a sequence-specific signal. DNA ligase is used to ligate the probes together, and, after washing, signal capture allows identification of the central base(s). Cleavage is performed after the central bases to remove the fluorescent tag and the last few bases. The process is then repeated for several more cycles, thereby identifying every "nth" base (for octamer with two specific central bases, the sequence would be known for positions 4, 5, 9, 10, 14, 15, etc.). Then, the double-stranded product is denatured, and the process begins anew with an initial probe that is one base shifted, so as to now read the sequence of the adjacent bases from the first round. This iterative process ensures that every base is read twice, once by two different probes.

SOLiD has short reads of less than 40 bases and thus requires extensive sequence alignment to reconstruct the genome. The "double reading" decreases the error rate. It is unique in using DNA ligase instead of polymerase, and some have reported difficulty in reading palindromic sequences [18].

Data Analysis

NGS technology has reached the point where sequencing is no longer the expensive, or time-consuming, or difficult part of the process; that dubious honor goes to the processing and interpretation of the data. An NGS run on one of the platforms described above can be done in a day or two at a reagent cost of no more than a few hundred dollars. Bioinformatics is a precious resource at most institutions, and NGS assays use much of it. In addition, NGS creates a massive amount of data (the human genome written out in text is, in itself, an 8 GB file), and dedicated servers and computers much exist to store this information, let alone process it.

Once the sequencing run is complete, computer programs are used to take all the sequence fragments and align them to each other based on overlapping sequences ("contigs") and, if possible, to a reference genome. For example, if the sequence is of a new extremophilic bacterium found deep inside an Antarctic glacier, the alignment will have to rely solely on the fragments from the NGS run in question. However, if the sequence is of a patient with a possible constitutional genetic syndrome, the program can use the reference human genome to help guide the reassembly. As mentioned above, longer reads make for an easier task; highly repetitive sequences pose a particularly challenging task (how to align a 40-base read within a 3 kb highly repetitive sequence?), and NGS is not the best methodology for looking at repeat size.

Coverage (also known as depth) is an important concept for NGS and is related to the number of times a nucleotide is present in a sequence fragment (how often that base was read). Coverage for a particular locus can be defined as the number of

times it was read multiplied by the average read length divided by the total genome length. Thus, platforms with longer reads may have greater coverage. A sequence with 1× coverage likely has many more errors (and is less reliable from an accurate variant detection standpoint) than a sequence with 10–30× coverage, which is the recommended depth for most clinically relevant constitutional aberrations; deeper sequencing, defined as coverage >1000×, is recommended for somatic variant detection, because cancer specimens can harbor variants of variable (and often low) allelic frequencies [19].

If a reference genome is available, one can be selective about what to analyze: just the exome? just a subset of genes? Because the bench side of NGS continues to get cheaper and quicker, we have arrived at the point where it may be easier to sequence an exome and just interpret a panel of a few genes than it is to perform automated Sanger sequencing on those genes individually. We are likely fast approaching the day when every patient has their genome sequenced early in life, and physicians query portions of it as health events occur over his/her lifetime.

Applications

NGS has found great utility in identifying constitutional genetic abnormalities such as loss-of-function or missense mutations in protein-coding genes. As mentioned above, some groups restrict their analysis to specific panels (for epilepsy, for pharmacogenomics, etc.). Also, on some platforms, copy number can be inferred from the exome data and coverage depth, although this varies. NGS is also capable of detecting large deletions or translocations.

NGS is also being increasingly used in pediatric solid tumors and leukemias/lymphomas. Several different commercial panels and many different custom-designed panels are available to analyze the sequence of clinically relevant genes such as tumor suppressors, proto-oncogenes, and genes with diagnostic, therapeutic, or prognostic implications. Such data have found clinical utility in identifying patients with tumor predisposition syndromes or with “druggable” or actionable mutations. One advantage to this approach is that the data can be reexamined as our medical knowledge grows and new clinical options become available. Microsatellite analysis is also currently being used to select patients for immunotherapy, such as the recent FDA-approval of pembrolizumab for adult and pediatric patients with unresectable/metastatic, microsatellite instability-high or mismatch repair deficient solid tumors that have progressed following prior treatment and who have no satisfactory alternative treatment options. Microsatellite instability status can be assessed by next-generation sequencing of the tumor [20–22], and this testing approach allows for simultaneous evaluation of other relevant genes. In addition, tumor mutation burden (TMB), which corresponds to the number of somatic mutations per megabase of sequencing, is now starting to be evaluated as a potential biomarker of response to immunotherapy in patients with various cancer types [23–26]. It is