



NGS – NEXT GENERATION SEQUENCING

PRINCÍPIOS BÁSICOS

Carla Daniele de Godoy

Assessora Científica do Laboratório de Anatomia
Patológica do Hospital Sírio-Libanês

- Maio/2019 -

Histórico

- Sequenciamento químico – Maxam e Gilbert, 1977.
- Sequenciamento enzimático – Frederick Sanger, 1977.
- Sequenciamento Sanger: semi-automatização, 1986.

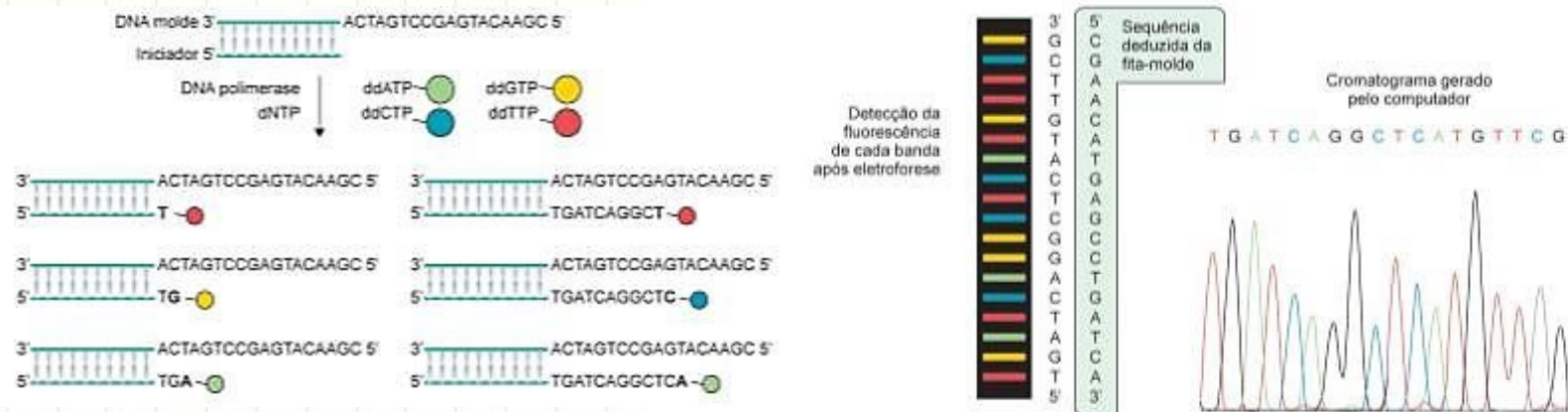


Figura 11.24 Sequenciamento de DNA pelo método de Sanger usando dideoxinucleotídeos fluorescentes.

Histórico

- Projeto Genoma Humano: iniciou em 1990. Primeiros dados publicados em 2001, genoma completo publicado em 2004, 20.500 genes identificados. Custo: \$3 bilhões.
- Em 2004, NHGRI criou programa de financiamento para reduzir o custo do sequenciamento do genoma humano para \$1.000 em 10 anos.

the human genome

International Human Genome Sequencing Consortium*

** A list of authors and their affiliations appears in the Supplementary Information*

The sequence of the human genome encodes the genetic instructions for human physiology, as well as rich information about human evolution. In 2001, the International Human Genome Sequencing Consortium reported a draft sequence of the euchromatic portion of the human genome. Since then, the international collaboration has worked to convert this draft into a genome sequence with high accuracy and nearly complete coverage. Here, we report the result of this finishing process. The current genome sequence (Build 35) contains 2.85 billion nucleotides interrupted by only 341 gaps. It covers ~99% of the euchromatic genome and is accurate to an error rate of ~1 event per 100,000 bases. Many of the remaining euchromatic gaps are associated with segmental duplications and will require focused work with new methods. The near-complete sequence, the first for a vertebrate, greatly improves the precision of biological analyses of the human genome including studies of gene number, birth and death. Notably, the human genome seems to encode only 20,000–25,000 protein-coding genes. The genome sequence reported here should serve as a firm foundation for biomedical research in the decades ahead.

NGS – Next Generation Sequencing

- Sequenciamento de nova geração = sequenciamento de segunda geração = Sequenciamento paralelo massivo de DNA.
- Melhorias na tecnologia de sequenciamento e métodos computacionais -> desenvolvimento de plataformas que diminuam o custo e tempo de análise de genoma e exoma.
- Aumentaram acurácia, eficiência, resolução e reprodutibilidade.
- Análise de várias amostras ao mesmo tempo e desenvolvimento de painéis genéticos para pesquisa de mutações.¹⁻³

1. Ross JS, Cronin M. Am J Clin Pathol. 2011 Oct;136(4):527-39.

2. Mardis ER. Annu Rev Genomics Hum Genet. 2008;9:387-402.

3. Metzker ML. Nature Reviews Genetics. 2009;11(1):31-46.

Plataformas

Table 1. Characteristics of current NGS platforms

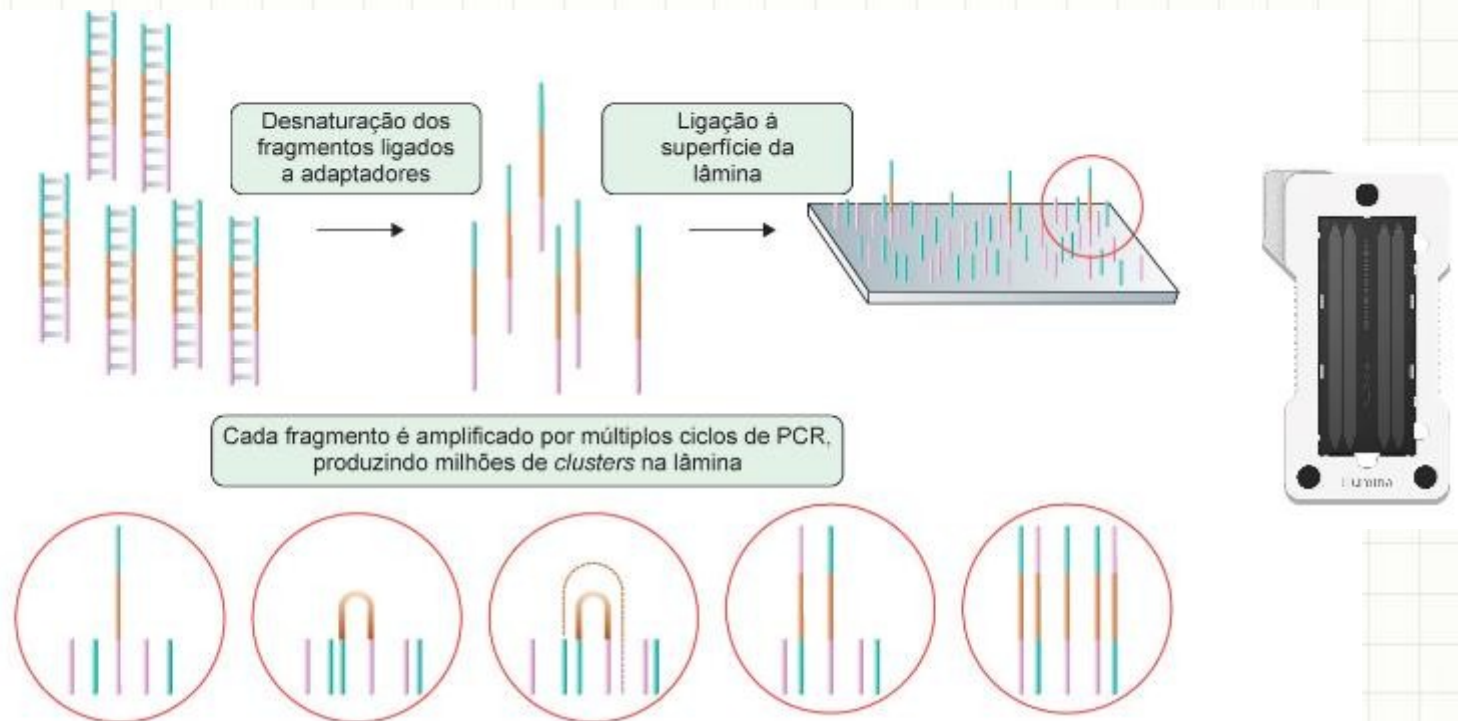
Plataform	Manufacturer	Sequencing chemistry	Detection	Read lenght, bp	Throughput	Run time	Dominant error type	Error Rate (%)
Miseq/MiseqDx	Illumina	Sequencing by synthesis-reversible terminator	Optical	1x36 to 2x300	1.7 - 7.5 Gb	4 - 39 h	Mismatch	0.1
Ion Torrent PGM	ThermoFisher	Sequencing by synthesis - hydrogen ion detection	Solid state	Up to 400	0.08 - 2 Gb	0.9 - 4.5 h	Indel	1
Ion Torrent S5/S5 XL	ThermoFisher	Sequencing by synthesis - hydrogen ion detection	Solid state	Up to 400	0.6 - 15 Gb	2.5 - 4 h	Indel	1
HiSeq 2500	Illumina	Sequencing by synthesis-reversible terminator	Optical	50 to 250	10 - 1.000 Gb	7h - 11 d	Mismatch	0.1
NextSeq 500	Illumina	Sequencing by synthesis-reversible terminator	Optical	1x75 to 2x150	32.5 - 120 Gb	11 - 16h	Mismatch	1
Ion Torrent Proton	ThermoFisher	Sequencing by synthesis - hydrogen ion detection	Solid state	Up to 200	10 - 15 Gb	2 - 4 h	Indel	1
HiSeq 3000/4000	ThermoFisher	Sequencing by synthesis - hydrogen ion detection	Solid state	2x150	650 - 1.500 Gb	< 1-3.5 d	Mismatch	0.1
HiSeq X	Illumina	Sequencing by synthesis-reversible terminator	Optical	2x150	1.6 - 1.8 Tb	< 3 d	Mismatch	0.1
GeneReader	QIAGEN	Sequencing by synthesis	Optical	Up to 150	2.5	35h	NA	NA
PacBio RS II	Pacific Biosciences	Single-molecule sequencing	Optical	Up to 60 kb	0,5 - 1 Gb	0.5 - 4 h	Indel	13
Nanopore	Nanopore	Single-molecule sequencing	Nanopore	Up to 100 kb	0,1 - 1 Gb	50 h	Indel/mismatch	12

Lefterova MI, *et al.* J Mol Diagn. 2015 Nov;17(6):623-34.

Desikan S, Narayanan S. Indian J Med Res. 2015 Jun;141(6):761-74.

Tipos de detecção

- Sequenciamento baseado em síntese
 - emissão de fluorescência (Illumina)



Tipos de detecção

- Sequenciamento baseado em síntese
 - emissão de fluorescência (Illumina)

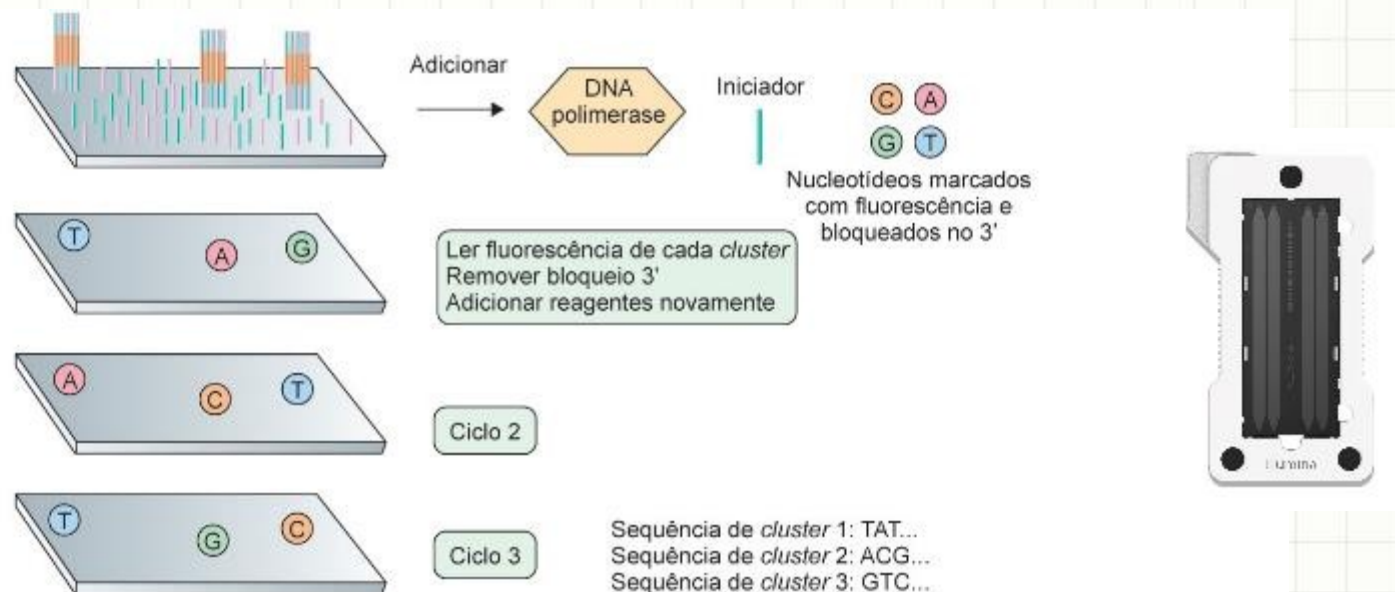
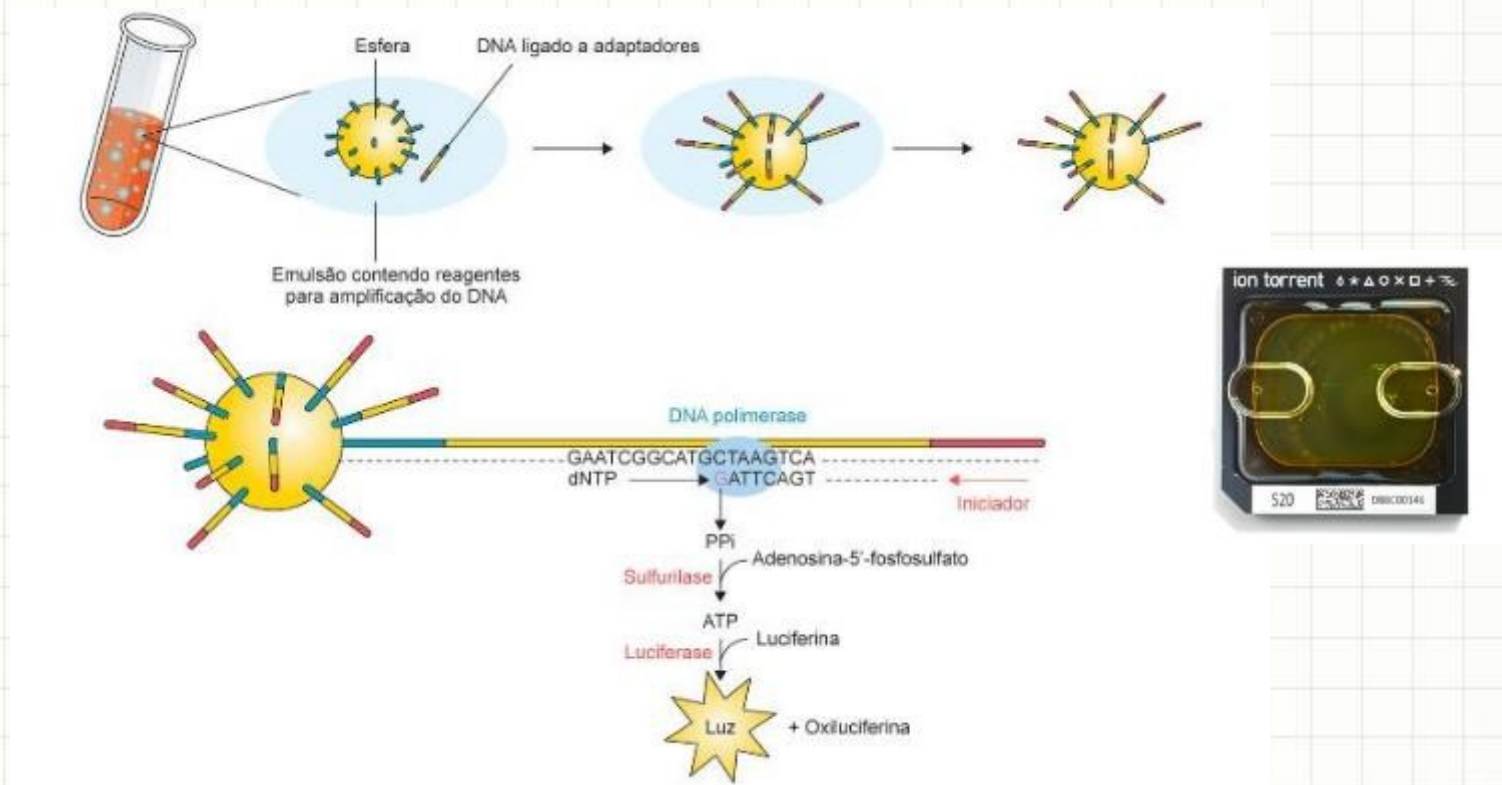


Figura 11.25 Sequenciamento de DNA na plataforma Illumina. Apenas três *clusters* são mostrados para simplificação.

Tipos de detecção

- Sequenciamento baseado em síntese
 - liberação de prótons de H^+ (Ion Torrent)



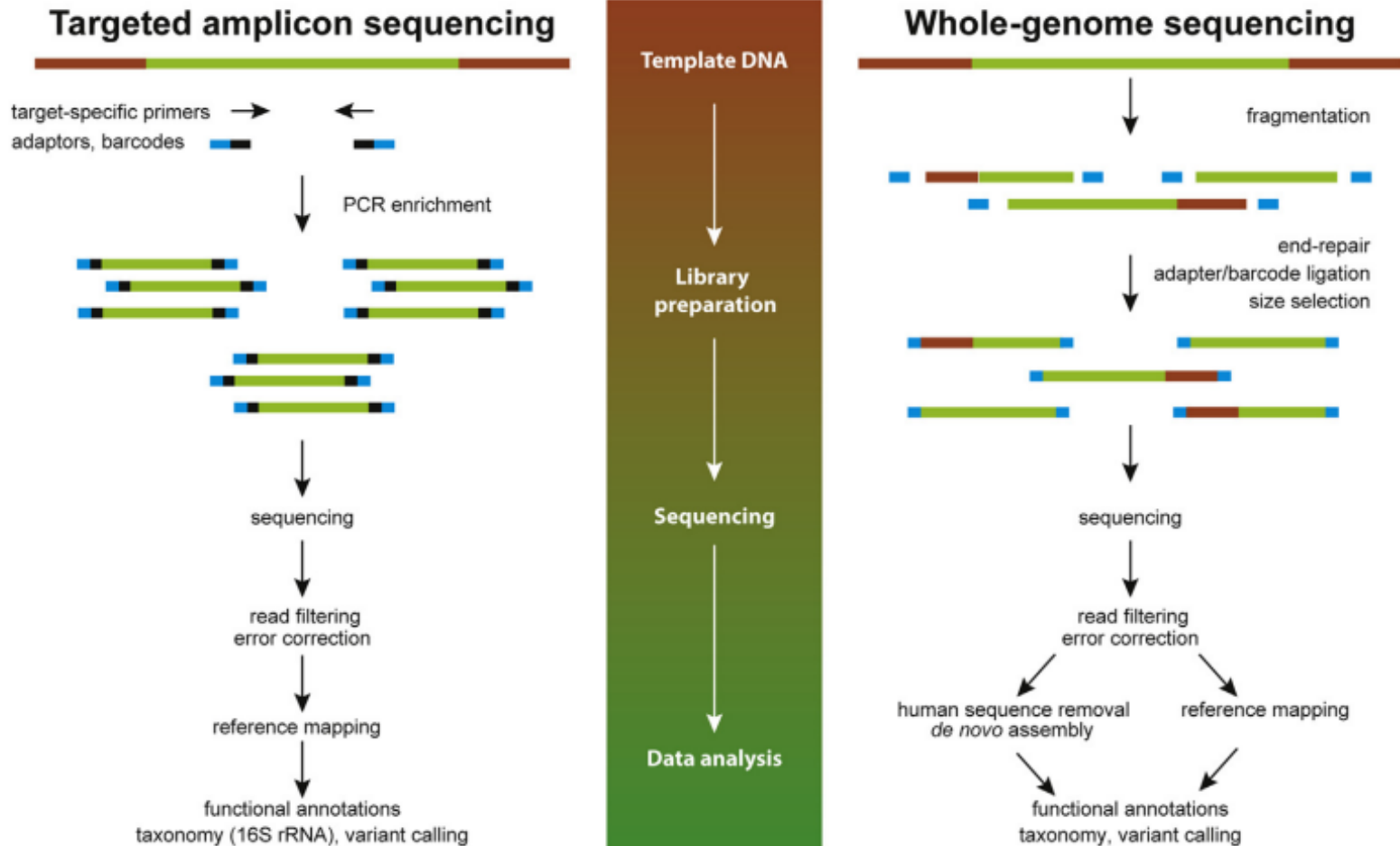
Análise dos dados

- Demultiplexing
- Métricas de qualidade do sequenciamento
- Mapeamento no genoma de referência
- Identificação de variantes
- Anotação de variantes
- Confecção de laudo

Tipos de abordagem

- WGS – Whole Genome Sequencing
- Exoma
- Transcriptoma/RNA-Seq
- DNA mitocondrial
- Metilação
- Painel alvo (captura ou amplicon)

Tipos de abordagem



Aplicabilidade

- Microbiologia:
 - Identificação de patógenos
 - Estudos de vigilância epidemiológica
 - Análises metagenômicas
 - Resistência a antibióticos e alterações na virulência
- Oncologia:
 - Painéis para detectar mutações somáticas
 - Painéis para detectar câncer hereditário (mutações germinativas)
 - Biópsia líquida (ctDNA)
- Outras aplicações:
 - Desordens metabólicas
 - Cardiologia
 - Hematologia
 - Screening em recém-nascidos (cell-free DNA)
 - Tipagem HLA
 - Análise forense



Oncologia

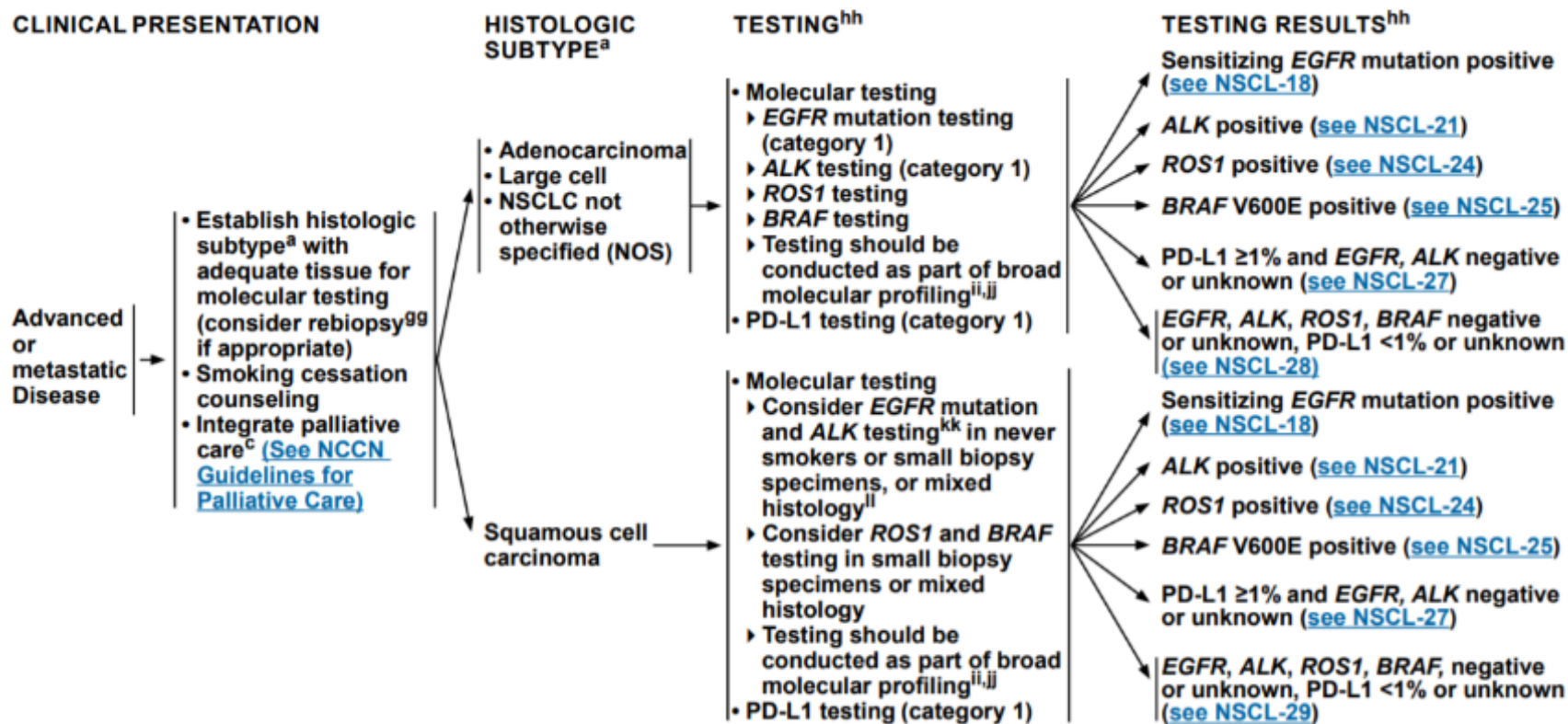
- Mais importante utilização do NGS na oncologia: painéis para detectar mutações (*hotspots* ou todos os éxons de um gene).
- Dado bruto -> alinhamento em relação à sequência referência -> chamada de variantes.
- Antes da chamada de variante: avaliação da performance do ensaio e qualidade dos dados.
- Variante identificada: reportar? Avaliar validade técnica e relevância clínica.

NGS – Next Generation Sequencing



NCCN Guidelines Version 4.2019 Non-Small Cell Lung Cancer

[NCCN Guidelines Index](#)
[Table of Contents](#)
[Discussion](#)



Terminologia

- Analista genômico: especialista (mestre ou doutor) que realiza análise, avalia a qualidade dos dados, utilidade clínica das variantes e cria o laudo. Laudo final será reportado em conjunto com um patologista e/ou equipe multidisciplinar.
- Q score: 0-20, probabilidade que a base sequenciada está correta (Q score 20 $\rightarrow p = 0,01$).
- QUAL (Escore de qualidade da variante): escore em log gerado durante a chamada de variante. Ex: score 90 $\rightarrow p = 1 \times 10^{-9}$. Varia de acordo com a plataforma do ensaio, método de captura e software de chamada de variante.
- Profundidade ou cobertura: número de vezes que determinada sequência foi sequenciada.

Terminologia

- Variante: variação encontrada em relação à sequência referência. O objetivo é identificar se a variante é proveniente de uma mutação somática.
- Mutações: evento que resulta em modificações no DNA genômico.
- Número de reads de variante: número de sequências independentes que apresentaram determinada variante (< 5 reads de variante = FP).
- Frequência do alelo variante: (número de reads de uma variante/cobertura daquela região) x 100.

Validade técnica

- Testes NGS aprovados pelo FDA.
 - MSK-IMPACT (Nov/2017)
 - Oncomine – Thermo (Jun/2017)
 - FoundationOne CDx (Nov/2017)
 - TSO500 – Illumina (em breve)
- Laboratórios desenvolvem seus próprios testes: licenciados pelo CLIA, Departamento de Saúde de Estado de NY, CAP.

npj | Genomic Medicine

www.nature.com/npjgenmed

COMMENT **OPEN**

FDA guidance for next generation sequencing-based testing:
balancing regulation and innovation in precision medicine

Frank Luh ¹ and Yun Yen²

npj Genomic Medicine (2018)3:28 ; doi:10.1038/s41525-018-0067-2

Validade técnica

- Método ortogonal confirmatório.
- Sanger não é a metodologia ideal para todas as variantes devido ao VAF de 10%-20%.
- Revisão manual do alinhamento reduz variantes FP e resultados incorretos. Software não substitui o analista no reconhecimento de padrões (principalmente em inserções e deleções).
- Estabelecer filtros, cutoffs, análises confirmatórias e utilização de amostras referências (DNA ou FFPE) durante validação (Acrometrix, Seracare e Horizon).

Validade técnica

- Mutações:
 - sensibilidade X resistência
 - somática X germinativa (ExAC e dbSNP – VAF 50% ou 100% potencialmente germinativas).
- Tipo de tumor: anotação da variante e interpretação (Clinvar, Cosmic, TCGA, MCG, cBioportal).
- VAF e celularidade tumoral: se discordantes, a variante pode ser inválida ou não reportável.
 - VAF maior que o esperado: mutação germinativa ou LoH.
 - VAF menor que o esperado: FP ou baixo nível de heterogeneidade tumoral (ex: mutações de resistência).

Validade técnica

- Mutações em hotspots recorrentes: devem ser reportadas mesmo que estejam no *borderline* da qualidade técnica (Ex: BRAF V600E, KRAS códon 12/13).
- Tipos de mutação clinicamente significativas: software pode classificar como VUS. Analista deverá determinar se a variante tem significado clínico, principalmente indels e mutações de perda de função
- VUS: laboratório determina sua “política” para reportar.
- Variantes de Síndrome de Câncer Hereditário: são consideradas achados incidentais e *screening* só é recomendado para pacientes alto risco.

Conclusões

- Extensa validação para assegurar confiabilidade, precisão e sensibilidade.
- Equipe técnica e analista genômico treinados para reconhecer possíveis falhas no processo de construção de biblioteca que poderá afetar os dados gerados.
- Reconhecer limitações do NGS: baixo valor preditivo negativo (não difere neoplasias benignas e malignas), regiões difíceis de sequenciar, pipelines de análise ainda não resolvem todos os problemas.
- Sequenciamento total do genoma/exoma tem uso limitado na prática clínica, uso de painéis tem maior relevância.