

Mutação

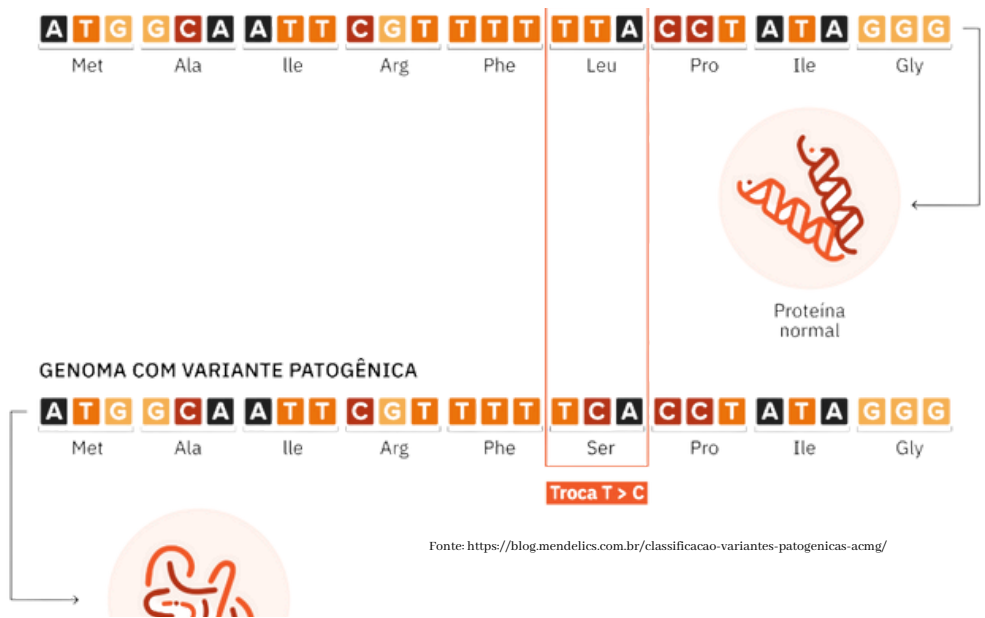


Variantes genéticas são alterações na sequência de DNA de um indivíduo, em relação a um genoma de referência. Elas são a principal fonte de diversidade entre as pessoas e podem ou não ter um impacto na saúde.

Polimorfismo

Tipos de Variantes:

- Variante de nucleotídeo único (SNV – Single nucleotide variants)
- Polimorfismo de nucleotídeo único (SNP)
- Variantes Indels
- Variação do número de cópias (CNV – Copy number variation)



Fonte: <https://blog.mendelics.com.br/classificacao-variantes-patogenicas-acmg/>



CHU, Duan; WEI, Lai. Nonsynonymous, synonymous and nonsense mutations in human cancer-related genes undergo stronger purifying selections than expectation. BMC cancer, v. 19, n. 1, p. 359, 2019.

LIN, Brian C. et al. In silico methods for predicting functional synonymous variants. Genome Biology, v. 24, n. 1, p. 126, 2023.

SHEN, Hui et al. Comprehensive characterization of human genome variation by high coverage whole-genome sequencing of forty four Caucasians. PloS one, v. 8, n. 4, p. e59494, 2013.

STERGACHIS, Andrew B. et al. Exonic transcription factor binding directs codon choice and affects protein evolution. Science, v. 342, n. 6164, p. 1367-1372, 2013.


VENTER, J. Craig et al. The sequence of the human genome. science, v. 291, n. 5507, p. 1304-1351, 2001.

As variantes são classificadas de acordo com as diretrizes estabelecidas pelo Colégio Americano de Genética Médica (ACMG). Essas diretrizes possuem 28 critérios diferentes aos quais são atribuídos diferentes pesos, em função das evidências disponíveis sobre a variante em questão.

- Variante benigna;
- Variante provavelmente benigna;
- Variante de significado incerto (VUS);
- Variante provavelmente patogênica;
- Variante patogênica.

ACMG Standards and Guidelines | Published: 05 March 2015

Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology

[Sue Richards PhD](#) , [Nazneen Aziz PhD](#), [Sherri Bale PhD](#), [David Bick MD](#), [Soma Das PhD](#), [Julie Gastier-Foster PhD](#), [Wayne W. Grody MD, PhD](#), [Madhuri Hegde PhD](#), [Elaine Lyon PhD](#), [Elaine Spector PhD](#), [Karl Voelkerding MD](#) & [Heidi L. Rehm PhD](#) on behalf of ; on behalf of the ACMG Laboratory Quality Assurance Committee

ARTICLE

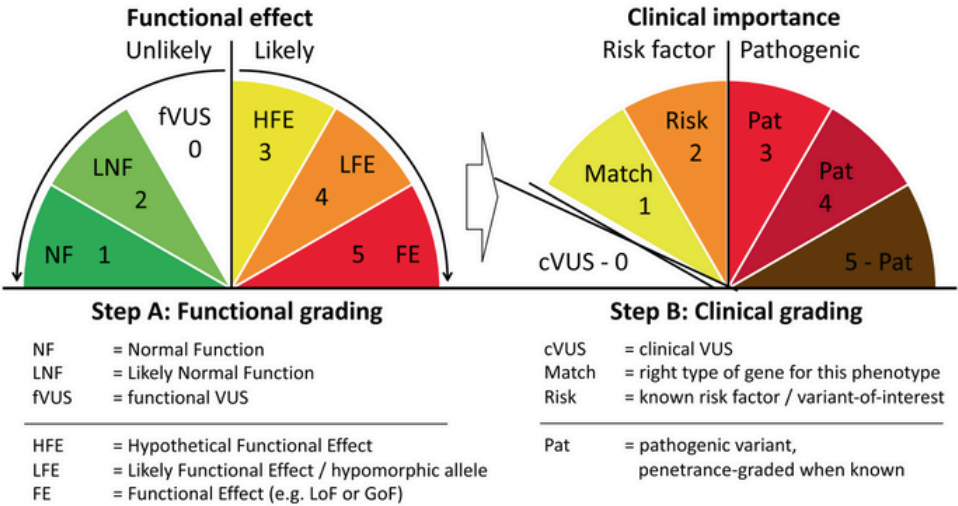


Stepwise ABC system for classification of any type of genetic variant

Gunnar Houge ¹ · Andreas Laner² · Sebahattin Cirak³ · Nicole de Leeuw⁴ · Hans Scheffer⁴ · Johan T. den Dunnen ⁵

Received: 22 July 2020 / Revised: 13 February 2021 / Accepted: 22 April 2021 / Published online: 13 May 2021
© The Author(s) 2021. This article is published with open access

Fig. 1 Graphical illustration of step A and B grading. The two scoring steps are functional (a) and clinical (b). Step A can be viewed as a likelihood scale from normal function to abnormal or no function, the balance being between fVUS and HFE. Step B grades from unknown clinical relevance (cVUS) to pathogenic with high penetrance, “risk” comprising both know risk factor genes and variants of interest (VOI).



Achados Secundários

Green RC, Berg JS, Grody WW, et al. ACMG recommendations for reporting of incidental findings in clinical exome and genome sequencing. Genet Med. 2013;15(7):565-574. doi:10.1038/gim.2013.73

Sequenciamento Completo do Exoma

Controle de qualidade dos dados brutos (FASTQ):

- Avaliar qualidade, remover adaptadores e reads ruins.

Alinhamento das sequências ao genoma de referência (.sam, BAM):

- Mapear reads ao genoma humano;
- Processamento do alinhamento;
- Ordenação do BAM, remoção de duplicatas e recalibragem de qualidade.

Chamada de variantes (VCF)

- Identificação de SNVs e indels.
- Filtragem de variantes
- Remover chamadas de baixa qualidade.

Anotação de variantes

- Associar variantes a genes, funções e bancos de dados clínicos.
- Validação e relatório clínico
- Confirmar variantes relevantes e emitir laudo.



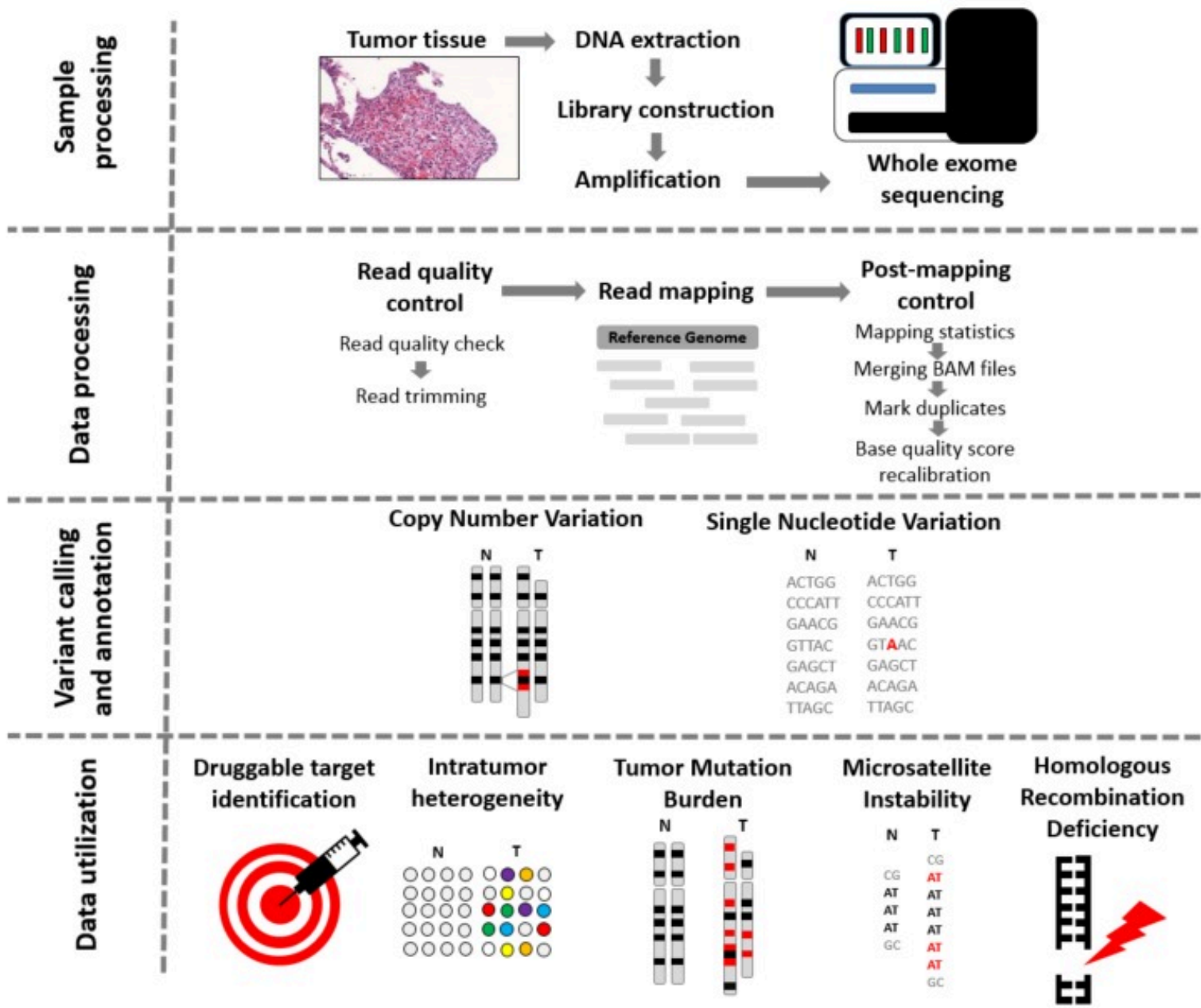
1000 Genomes Project

General Background

In 2008, the international 1000 Genomes Project Consortium launched the [1000 Genomes Project](#) to develop a public resource of human genetic variation. This resource supports genome-wide association studies and other studies involving genetic variation, health and disease. All data for the 1000 Genomes Project are freely available to the public through the [1000 Genomes Project](#) website.

ClinVar

ClinVar aggregates information about genomic variation and its relationship to human health.



BARTHA, Áron; GYÖRFFY, Balázs. Comprehensive outline of whole exome sequencing data analysis tools available in clinical oncology. Cancers, v. 11, n. 11, p. 1725, 2019.
BAGGER, Frederik Otzen et al. Whole genome sequencing in clinical practice. BMC medical genomics, v. 17, n. 1, p. 59, 2024.
CAUSEY, Jason L. et al. DNAP: a pipeline for DNA-seq data analysis. Scientific reports, v. 8, n. 1, p. 6793, 2018.
KAMPS, Rick et al. Next-generation sequencing in oncology: genetic diagnosis, risk prediction and cancer classification. International journal of molecular sciences, v. 18, n. 2, p. 308, 2017.

wANNOVAR

ANNOVAR is a rapid, efficient tool to annotate functional consequences of genetic variation from high-throughput sequencing data. wANNOVAR provides easy and intuitive web-based access to the most popular functionalities of the ANNOVAR software

Ferramentas bioinformáticas para a chamada de variantes em dados de sequenciamento de DNA humano

The Genome Analysis Toolkit (GATK): (MCKENNA, Aaron et al., 2010)

- Baseado na programação funcional MapReduce;
- Suporta arquivos BAM com alinhamentos emitidos pela maioria dos alinhadores de NGS e testado com diversos BAMs gerados por várias ferramentas de alinhamento públicas;
- Fornece múltiplas abordagens para paralelização de tarefas;
- Suporte a grandes volumes de dados;
- Permite excluir leituras de baixa qualidade ou regiões problemáticas, garantindo resultados mais confiáveis.

FreeBayes (Garrison E, Marth G, 2012)

- Detector de variantes genéticas bayesiano projetado para encontrar pequenos polimorfismos, especificamente SNPs (polimorfismos de nucleotídeo único), indels (inserções e deleções), MNPs (polimorfismos de nucleotídeos múltiplos) e eventos complexos (eventos compostos de inserção e substituição) menores que o comprimento de um alinhamento de sequenciamento de leitura curta;
- Usar um conjunto de variantes de entrada (VCF) como informação prévia para guiar a chamada de variantes;
- Precisão menor em variantes de baixa frequência

DeepVariant

- Ferramenta de chamada de variantes baseada em aprendizado profundo, que recebe leituras alinhadas (em formato BAM ou CRAM), produz tensores de imagem de empilhamento a partir delas, classifica cada tensor usando uma rede neural convolucional e, finalmente, relata os resultados em um arquivo VCF ou gVCF padrão;
- Chamada de variantes somáticas;
- Alta precisão, flexibilidade, velocidade e custo-benefício.

VCF (Variant Call Format)

VCF é um formato de arquivo de texto padronizado para armazenar informações sobre variações genéticas (SNPs, Indels e outras variantes estruturais). Ele contém linhas de metainformação, uma linha de cabeçalho e, em seguida, linhas de dados, cada uma contendo informações sobre uma posição no genoma. O formato também pode conter informações de genótipo em amostras para cada posição.

Estrutura:

- Metainformação;
- Linha de cabeçalho;
- Dados das variantes (uma linha por variante).

Utilização na Análise de Variantes:

- Compartilhamento e reprodutibilidade;
- Interoperabilidade entre as plataormas;
- Integração com bancos de dados;
- Filtragem, anotação funcional e comparação de variantes;
- Interpretação clínica e de pesquisa.

```
##fileformat=VCFv4.1
##fileDate=20090805
##source=myImputationProgramV3.1
##reference=file:///seq/references/1000GenomesPilot-NCBI36.fasta
##contig=<ID=20,length=62435964,assembly=B36,md5=f126cdf8a6e0c7f379d618ff66beb2da,species="Homo sapiens",taxonomy=x>
##phasing=partial
##INFO=<ID=NS,Number=1,Type=Integer,Description="Number of Samples With Data">
##INFO=<ID=DP,Number=1,Type=Integer,Description="Total Depth">
##INFO=<ID=AF,Number=A,Type=Float,Description="Allele Frequency">
##INFO=<ID=AA,Number=1,Type=String,Description="Ancestral Allele">
##INFO=<ID=DB,Number=0,Type=Flag,Description="dbSNP membership, build 129">
##INFO=<ID=H2,Number=0,Type=Flag,Description="HapMap2 membership">
##FILTER=<ID=q10,Description="Quality below 10">
##FILTER=<ID=s50,Description="Less than 50% of samples have data">
##FORMAT=<ID=GT,Number=1,Type=String,Description="Genotype">
##FORMAT=<ID=GQ,Number=1,Type=Integer,Description="Genotype Quality">
##FORMAT=<ID=DP,Number=1,Type=Integer,Description="Read Depth">
##FORMAT=<ID=HQ,Number=2,Type=Integer,Description="Haplotype Quality">
#CHROM POS ID REF ALT QUAL FILTER INFO FORMAT NA00001 NA00002 NA00003
20 14370 rs6054257 G A 29 PASS NS=3;DP=14;AF=0.5;DB;H2 GT:GQ:DP:HQ 0|0:48:1:51,51 1|0:48:8:51,51 1/1:43:5:.,.
20 17330 . T A 3 q10 NS=3;DP=11;AF=0.017 GT:GQ:DP:HQ 0|0:49:3:58,50 0|1:3:5:65,3 0/0:41:3
20 1110696 rs6040355 A G,T 67 PASS NS=2;DP=10;AF=0.333,0.667;AA=T;DB GT:GQ:DP:HQ 1|2:21:6:23,27 2|1:2:0:18,2 2/2:35:4
20 1230237 . T . 47 PASS NS=3;DP=13;AA=T GT:GQ:DP:HQ 0|0:54:7:56,60 0|0:48:4:51,51 0/0:61:2
20 1234567 microsat1 GTC G,GTCT 50 PASS NS=3;DP=9;AA=G GT:GQ:DP 0/1:35:4 0/2:17:2 1/1:40:3
```