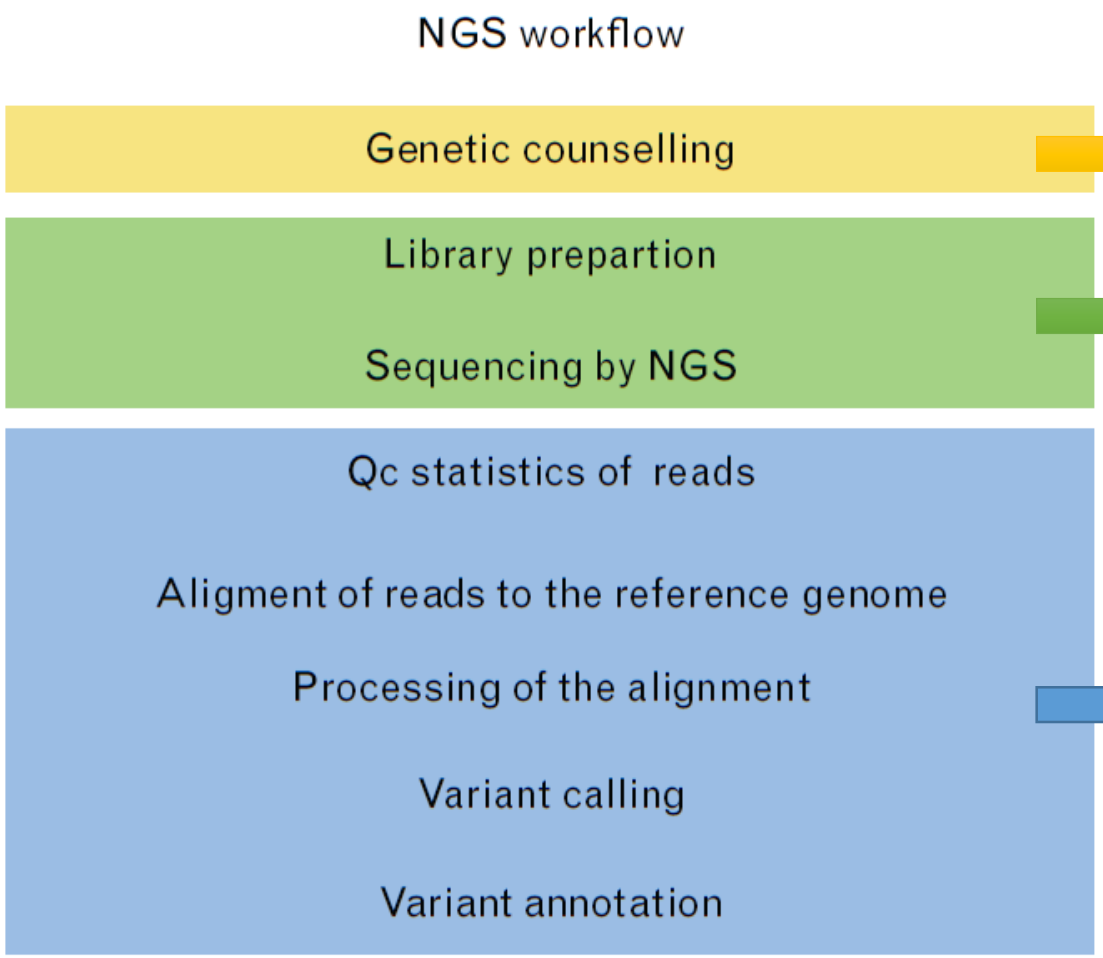
Школа анализа NGS данных MGNGS School’19
Москва, 28.10—01.11.2019

Правила интерпретации NGS данных

А.В. Марахонов, к.б.н.,
с.н.с. ФГБНУ «МГНЦ»

NGS workflow

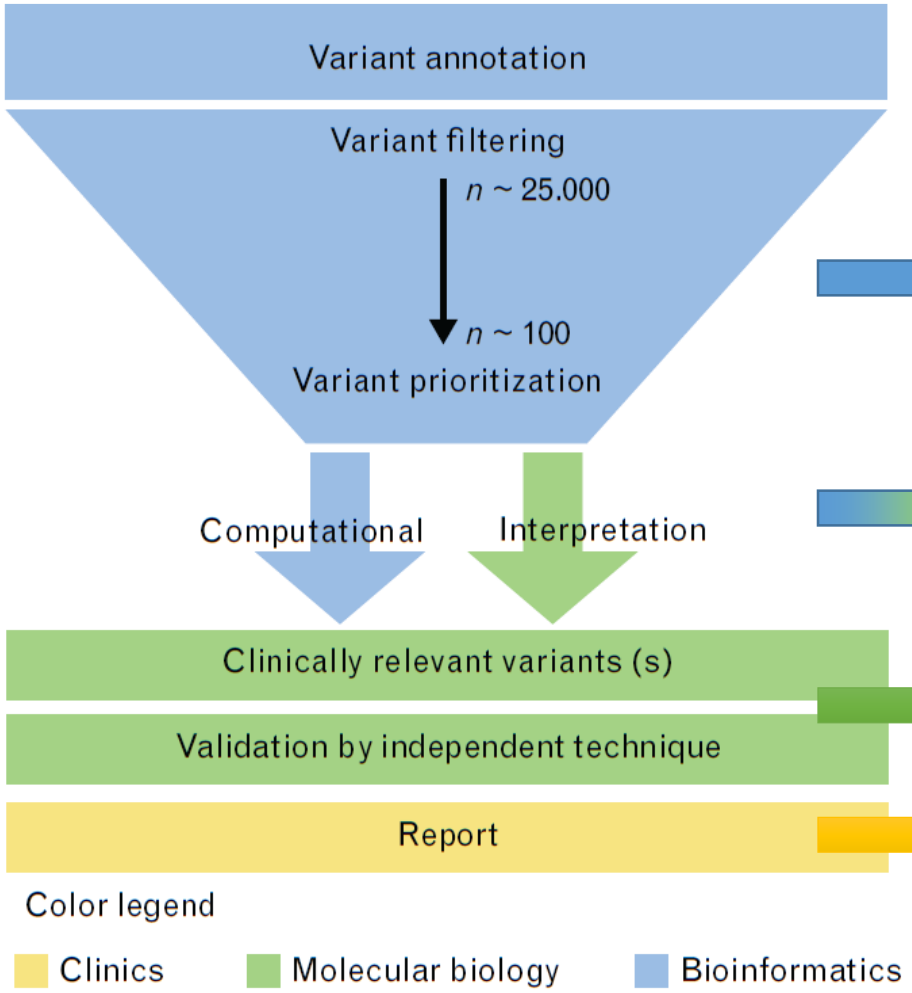


Формирование клинической гипотезы, направление на NGS

Лабораторная часть, «мокрая» биология

Биоинформатическая часть

NGS workflow



Фильтрация вариантов

Интерпретация вариантов

Выбор релевантного варианта(ов), подтверждение

Заключение о результатах секвенирования

Критерии ACMG

ACMG STANDARDS AND GUIDELINES | Genetics inMedicine

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[1], Nazneen Aziz, PhD[2,16], Sherri Bale, PhD[3], David Bick, MD[4], Soma Das, PhD[5], Julie Gastier-Foster, PhD[6,7,8], Wayne W. Grody, MD, PhD[9,10,11], Madhuri Hegde, PhD[12], Elaine Lyon, PhD[13], Elaine Spector, PhD[14], Karl Voelkerding, MD[13] and Heidi L. Rehm, PhD[15]; on behalf of the ACMG Laboratory Quality Assurance Committee

Руководство по интерпретации данных последовательности ДНК человека, полученных методами массового параллельного секвенирования (MPS) (редакция 2018, версия 2)

Рыжкова О.П.[1], Кардымон О.Л.[1], Прохорчук Е.Б.[2], Коновалов Ф.А.[3], Масленников А.Б.[4], Степанов В.А.[5], Афанасьев А.А.[6], Заклязьминская Е.В.[7], Ребриков Д.В.[8], Савостьянов К.В.[9], Глотов А.С.[10,11], Костарева А.А.[12], Павлов А.Е.[13], Голубенко М.В.[5], Поляков А.В.[1], Куцев С.И.[1]

Критерии ACMG

Population data

_ PS4 The prevalence of the variant in affected individuals is significantly increased compared with the prevalence in controls

_ PM2 Absent from controls (or at extremely low frequency if recessive) in Exome Sequencing Project, 1000 Genomes Project, or Exome Aggregation Consortium

_ BS1 Allele frequency is greater than expected for d

_ BS2 Observed in a healthy adult individual for a re with full penetrance expected at an early age

_ BA1 Allele frequency is >5% in Exome Sequencing P

Computational and predictive data

_ PVS1 null variant (nonsense, frameshift, canonical is a known mechanism of disease

_ PS1 Same amino acid change as a previously estab

_ PM4 Protein length changes as a result of in-frame

_ PM5 Novel missense change at an amino acid resid before

_ PP3 Multiple lines of computational evidence supp splicing impact, etc.)

_ BP1 Missense variant in a gene for which primarily

_ BP3 In-frame deletions/insertions in a repetitive re

_ BP4 Multiple lines of computational evidence sugg etc.)

_ BP7 A synonymous (silent) variant for which splici creation of a new splice site AND the nucleotide is no

Functional data

_ PS3 Well-established in vitro or in vivo functional studies supportive of a damaging effect on the gene or gene product

_ PM1 Located in a mutational hot spot and/or critical and well-established functional domain (e.g., active site of an enzyme) without benign variation

_ PP2 Missense variant in a gene that has a low rate of benign missense variation and in which missense variants are a common mechanism of disease

_ BS3 Well-established in vitro or in vivo functional studies show no damaging effect on protein function or splicing

Segregation data

_ PP1 (Strong evidence) Cosegregation with disease in multiple affected family members in a gene definitively known to cause the disease

_ PP1 (Moderate evidence) Cosegregation with disease in multiple affected family members in a gene definitively known to cause the disease

_ PP1 Cosegregation with disease in multiple affected family members in a gene definitively known to cause the disease

_ BS4 Lack of segregation in affected members of a family

De novo data

_ PS2 De novo (both mat

_ PM6 Assumed de novo

Allelic data

_ PM3 For recessive diso

_ BP2 Observed in trans variant in any inheritan

Other database data

_ PP5 Reputable source recently reports variant as pathogenic, but the evidence is not available to the laboratory to perform an independent evaluation

_ BP6 Reputable source recently reports variant as benign, but the evidence is not available to the laboratory to perform an independent evaluation

Other data

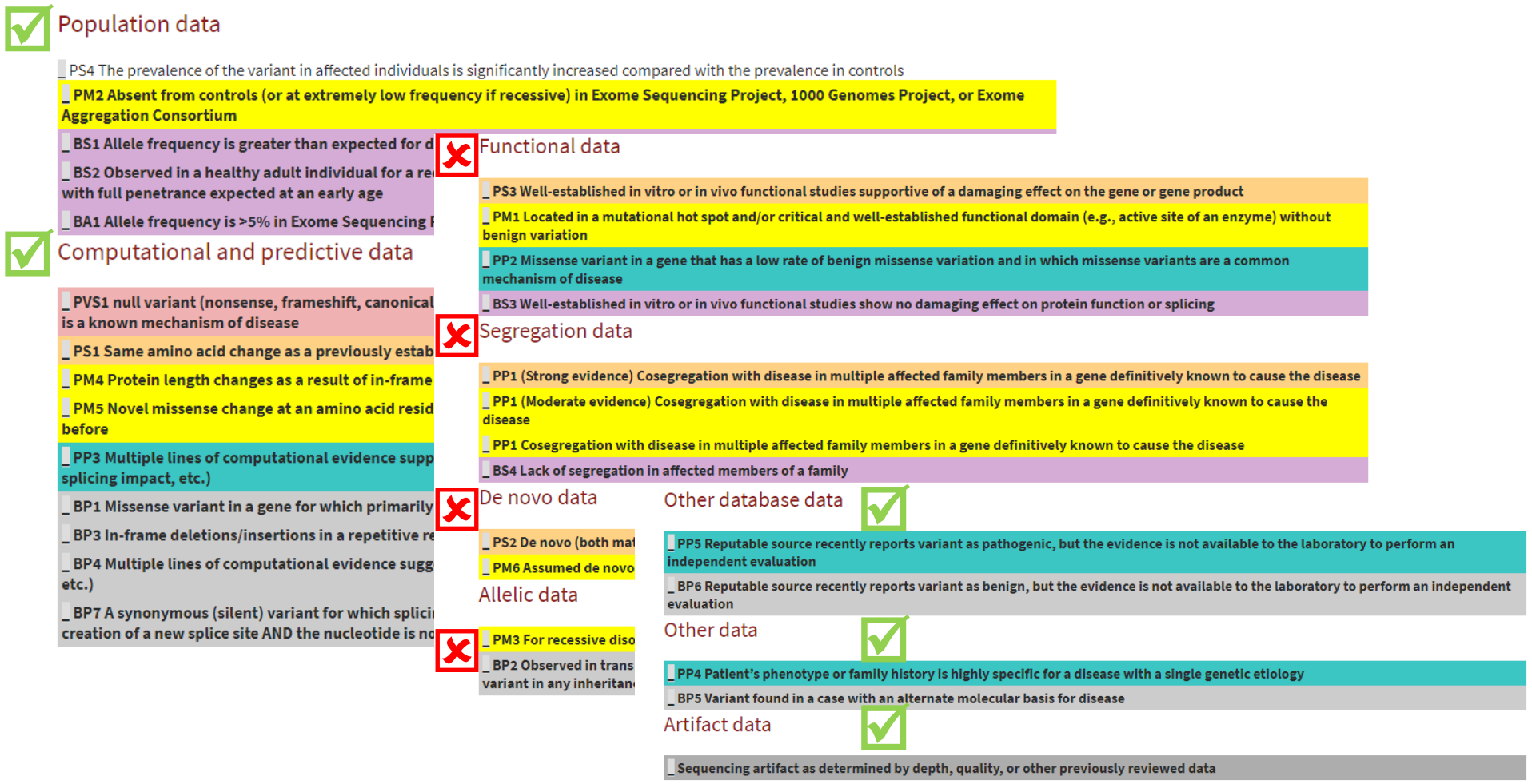
_ PP4 Patient's phenotype or family history is highly specific for a disease with a single genetic etiology

_ BP5 Variant found in a case with an alternate molecular basis for disease

Artifact data

_ Sequencing artifact as determined by depth, quality, or other previously reviewed data

Критерии ACMG



Применение критериев

Evidence of pathogenicity	Category
Very strong	PVS1 null variant (nonsense, frameshift, canonical ±1 or 2 splice sites, initiation codon, single or multiexon deletion) in a gene where LOF is a known mechanism of disease Caveats: • Beware of genes where LOF is not a known disease mechanism (e.g., *GFAP*, *MYH7*) • Use caution interpreting LOF variants at the extreme 3′ end of a gene • Use caution with splice variants that are predicted to lead to exon skipping but leave the remainder of the protein intact • Use caution in the presence of multiple transcripts
Strong	PS1 Same amino acid change as a previously established pathogenic variant regardless of nucleotide change Example: Val→Leu caused by either G>C or G>T in the same codon Caveat: Beware of changes that impact splicing rather than at the amino acid/protein level PS2 De novo (both maternity and paternity confirmed) in a patient with the disease and no family history Note: Confirmation of paternity only is insufficient. Egg donation, surrogate motherhood, errors in embryo transfer, and so on, can contribute to nonmaternity. PS3 Well-established in vitro or in vivo functional studies supportive of a damaging effect on the gene or gene product Note: Functional studies that have been validated and shown to be reproducible and robust in a clinical diagnostic laboratory setting are considered the most well established. PS4 The prevalence of the variant in affected individuals is significantly increased compared with the prevalence in controls Note 1: Relative risk or OR, as obtained from case–control studies, is >5.0, and the confidence interval around the estimate of relative risk or OR does not include 1.0. See the article for detailed guidance. Note 2: In instances of very rare variants where case–control studies may not reach statistical significance, the prior observation of the variant in multiple unrelated patients with the same phenotype, and its absence in controls, may be used as moderate level of evidence.

Применение критериев

Moderate	PM1 Located in a mutational hot spot and/or critical and well-established functional domain (e.g., active site of an enzyme) without benign variation
	PM2 Absent from controls (or at extremely low frequency if recessive) (**Table 6**) in Exome Sequencing Project, 1000 Genomes Project, or Exome Aggregation Consortium
	Caveat: Population data for insertions/deletions may be poorly called by next-generation sequencing.
	PM3 For recessive disorders, detected in *trans* with a pathogenic variant
	Note: This requires testing of parents (or offspring) to determine phase.
	PM4 Protein length changes as a result of in-frame deletions/insertions in a nonrepeat region or stop-loss variants
	PM5 Novel missense change at an amino acid residue where a different missense change determined to be pathogenic has been seen before
	Example: Arg156His is pathogenic; now you observe Arg156Cys
	Caveat: Beware of changes that impact splicing rather than at the amino acid/protein level.
	PM6 Assumed de novo, but without confirmation of paternity and maternity
Supporting	PP1 Cosegregation with disease in multiple affected family members in a gene definitively known to cause the disease
	Note: May be used as stronger evidence with increasing segregation data
	PP2 Missense variant in a gene that has a low rate of benign missense variation and in which missense variants are a common mechanism of disease
	PP3 Multiple lines of computational evidence support a deleterious effect on the gene or gene product (conservation, evolutionary, splicing impact, etc.)
	Caveat: Because many in silico algorithms use the same or very similar input for their predictions, each algorithm should not be counted as an independent criterion. PP3 can be used only once in any evaluation of a variant.
	PP4 Patient's phenotype or family history is highly specific for a disease with a single genetic etiology
	PP5 Reputable source recently reports variant as pathogenic, but the evidence is not available to the laboratory to perform an independent evaluation

LOF, loss of function; OR, odds ratio.

Применение критериев: доброкачественные

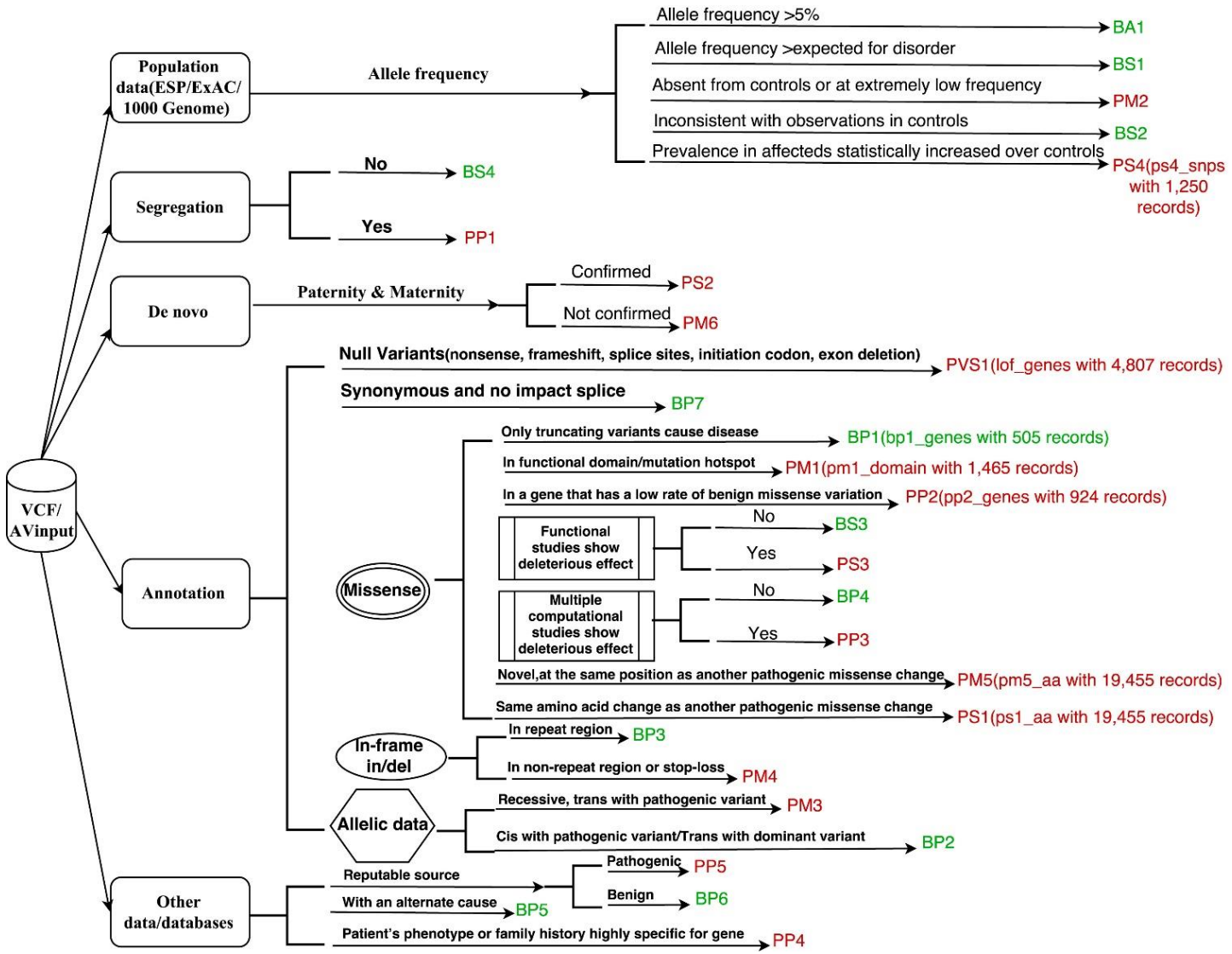
Evidence of benign impact	Category
Stand-alone	BA1 Allele frequency is >5% in Exome Sequencing Project, 1000 Genomes Project, or Exome Aggregation Consortium
Strong	BS1 Allele frequency is greater than expected for disorder (see **Table 6**)
	BS2 Observed in a healthy adult individual for a recessive (homozygous), dominant (heterozygous), or X-linked (hemizygous) disorder, with full penetrance expected at an early age
	BS3 Well-established in vitro or in vivo functional studies show no damaging effect on protein function or splicing
	BS4 Lack of segregation in affected members of a family
	Caveat: The presence of phenocopies for common phenotypes (i.e., cancer, epilepsy) can mimic lack of segregation among affected individuals. Also, families may have more than one pathogenic variant contributing to an autosomal dominant disorder, further confounding an apparent lack of segregation.
Supporting	BP1 Missense variant in a gene for which primarily truncating variants are known to cause disease
	BP2 Observed in *trans* with a pathogenic variant for a fully penetrant dominant gene/disorder or observed in *cis* with a pathogenic variant in any inheritance pattern
	BP3 In-frame deletions/insertions in a repetitive region without a known function
	BP4 Multiple lines of computational evidence suggest no impact on gene or gene product (conservation, evolutionary, splicing impact, etc.)
	Caveat: Because many in silico algorithms use the same or very similar input for their predictions, each algorithm cannot be counted as an independent criterion. BP4 can be used only once in any evaluation of a variant.
	BP5 Variant found in a case with an alternate molecular basis for disease
	BP6 Reputable source recently reports variant as benign, but the evidence is not available to the laboratory to perform an independent evaluation
	BP7 A synonymous (silent) variant for which splicing prediction algorithms predict no impact to the splice consensus sequence nor the creation of a new splice site AND the nucleotide is not highly conserved

Применение критериев

Pathogenic	(i) 1 Very strong (PVS1) *AND* (a) ≥1 Strong (PS1–PS4) *OR* (b) ≥2 Moderate (PM1–PM6) *OR* (c) 1 Moderate (PM1–PM6) and 1 supporting (PP1–PP5) *OR* (d) ≥2 Supporting (PP1–PP5) (ii) ≥2 Strong (PS1–PS4) *OR* (iii) 1 Strong (PS1–PS4) *AND* (a)≥3 Moderate (PM1–PM6) *OR* (b)2 Moderate (PM1–PM6) *AND* ≥2 Supporting (PP1–PP5) *OR* (c)1 Moderate (PM1–PM6) *AND* ≥4 supporting (PP1–PP5)
Likely pathogenic	(i) 1 Very strong (PVS1) *AND* 1 moderate (PM1–PM6) *OR* (ii) 1 Strong (PS1–PS4) *AND* 1–2 moderate (PM1–PM6) *OR* (iii) 1 Strong (PS1–PS4) *AND* ≥2 supporting (PP1–PP5) *OR* (iv) ≥3 Moderate (PM1–PM6) *OR* (v) 2 Moderate (PM1–PM6) *AND* ≥2 supporting (PP1–PP5) *OR* (vi) 1 Moderate (PM1–PM6) *AND* ≥4 supporting (PP1–PP5)

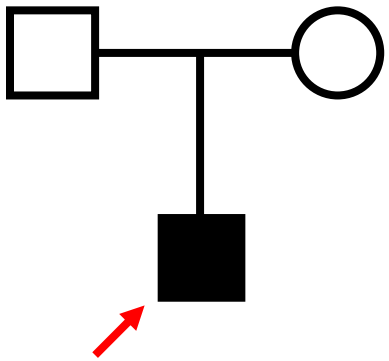
Benign	(i) 1 Stand-alone (BA1) *OR* (ii) ≥2 Strong (BS1–BS4)
Likely benign	(i) 1 Strong (BS1–BS4) and 1 supporting (BP1–BP7) *OR* (ii) ≥2 Supporting (BP1–BP7)
Uncertain significance	(i) Other criteria shown above are not met OR (ii) the criteria for benign and pathogenic are contradictory

Применение критериев



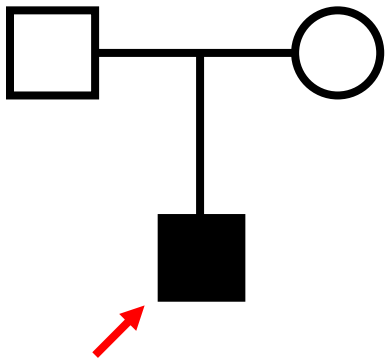
Фильтрация вариантов

- Аллельное состояние?



Фильтрация вариантов

- Аллельное состояние?



Homozygous/Compound heterozygous?
Heterozygous?
Hemizygous?

Фильтрация вариантов

- Присутствие в «контрольных» базах данных (dbSNP, Clinvar…)?

Фильтрация вариантов

- Присутствие в «контрольных» базах данных (dbSNP, Clinvar…)?

- CFTR:delF508 — rs1297060838
- PAH:R408W — rs5030858
- GJB2:35delG — rs80338939

Популяционная частота?

- Рекомендации РОМГ:

- PM2: вариант отсутствует в контрольных выборках (или встречается с крайне низкой частотой): для аутосомно-доминантных заболеваний частота аллеля не должна превышать 0,01%, для аутосомно-рецессивных заболеваний — 0,5%, для доминантных Х-сцеплённых — 0,03%, для рецессивных Х-сцеплённых — 0,5%.

Популяционная частота?

GJB2:35delG

Population Frequencies

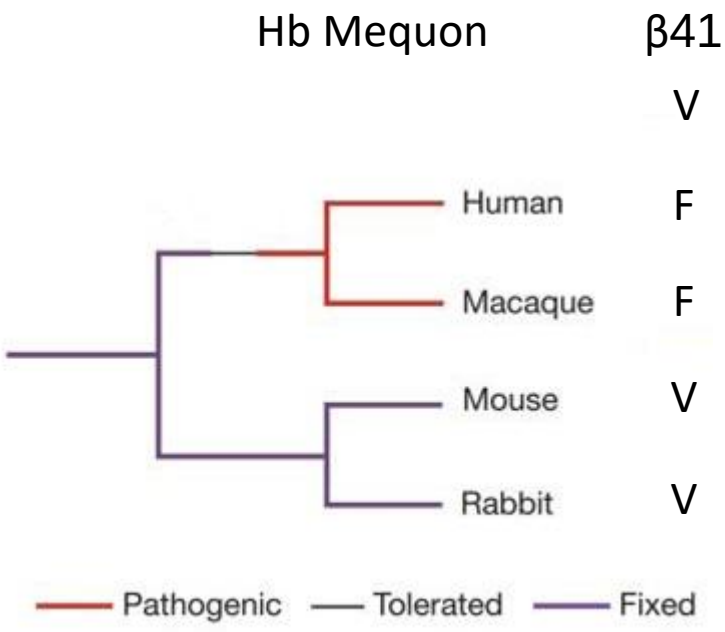
Population	Allele Count	Allele Number	Number of Homozygotes	Allele Frequency
European (non-Finnish)	1217	127068	4	0.009578
European (Finnish)	210	25124	6	0.008359
Other	55	7206	0	0.007633
Latino	169	35434	0	0.004769
Ashkenazi Jewish	35	10362	0	0.003378
African	26	24940	0	0.001043
South Asian	25	30612	0	0.0008167
East Asian	0	19950	0	0.000
Total	**1737**	**280696**	**10**	**0.006188**

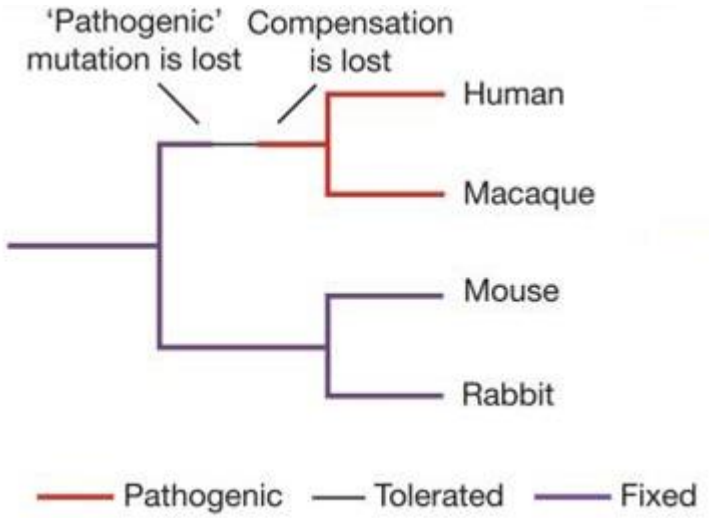
Консервативность?

- У приматов?
- У позвоночных?

10.1038/nature14497

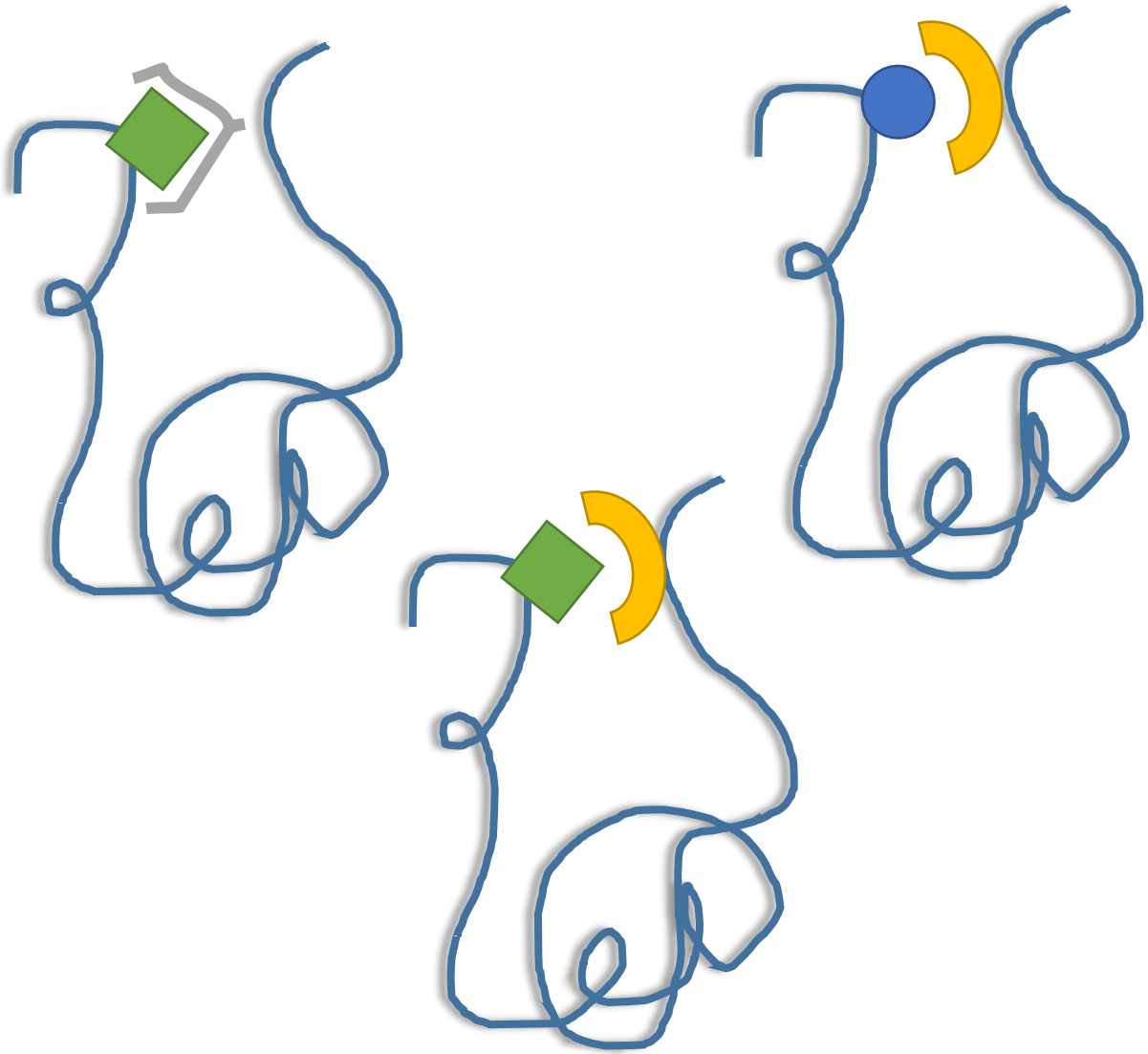
Консервативность?





10.1016/j.sbi.2017.10.009

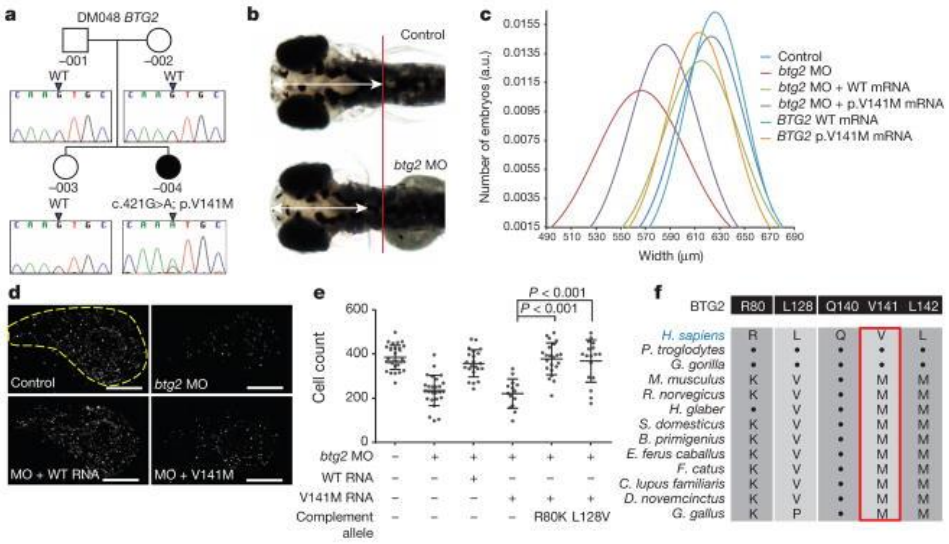
Консервативность?



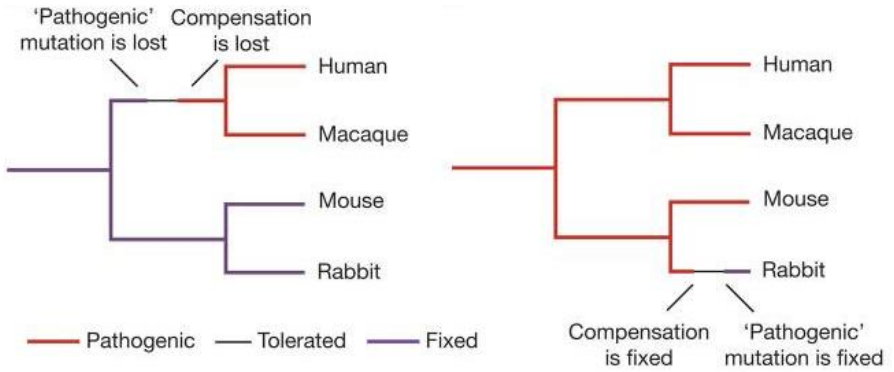
Консервативность?

De novo missense BTG2 (p.V141M)

dbNSFP3.5 predictions	
PolyPhen2 prediction	Benign
SIFT prediction	Tolerated
LRT prediction	Neutral
MutationTaster prediction	Polymorphism
MutationAssessor prediction	Neutral
FATHMM	Tolerated
fathmm-MKL	Neutral
M-CAP	Tolerated
CADD The larger the score the more likely the SNP is damaging (PHRED-like)	12.72
MetaSVM	Tolerated
MetaLR	Tolerated
PhyloP 20way The larger the score, the more conserved the site (max 1.199000).	0.165000
PhyloP 100way The larger the score, the more conserved the site (max 10.003000).	0.513000
GERP RS The larger the score, the more conserved the site (max 6.17).	2.37
ExAC	No data
1000 Genomes	No data
gnomAD	No data
Interpro domain	Anti-proliferative protein



cis-suppression



10.1038/nature14497