NGS interpretation and reporting: in the view of pathologists

Jun Kang

Catholic University of Korea

2022 02 12

Contents

- Allele frequency
- Variant pathogenecity

Questions

- Somatic vs germline
- False positive or false negative (cut-off)
- Driver vs passenger (ultra-hypermutation, POLE)
- Tier

Allele frequency

- BRCA1 mutation: Positive - p.Glu649Ter (c.1945G>T)
- Variant allele frequency 63.84%
- Tumor cell percentage: 70%
- BRCA negative in blood sample

Questions

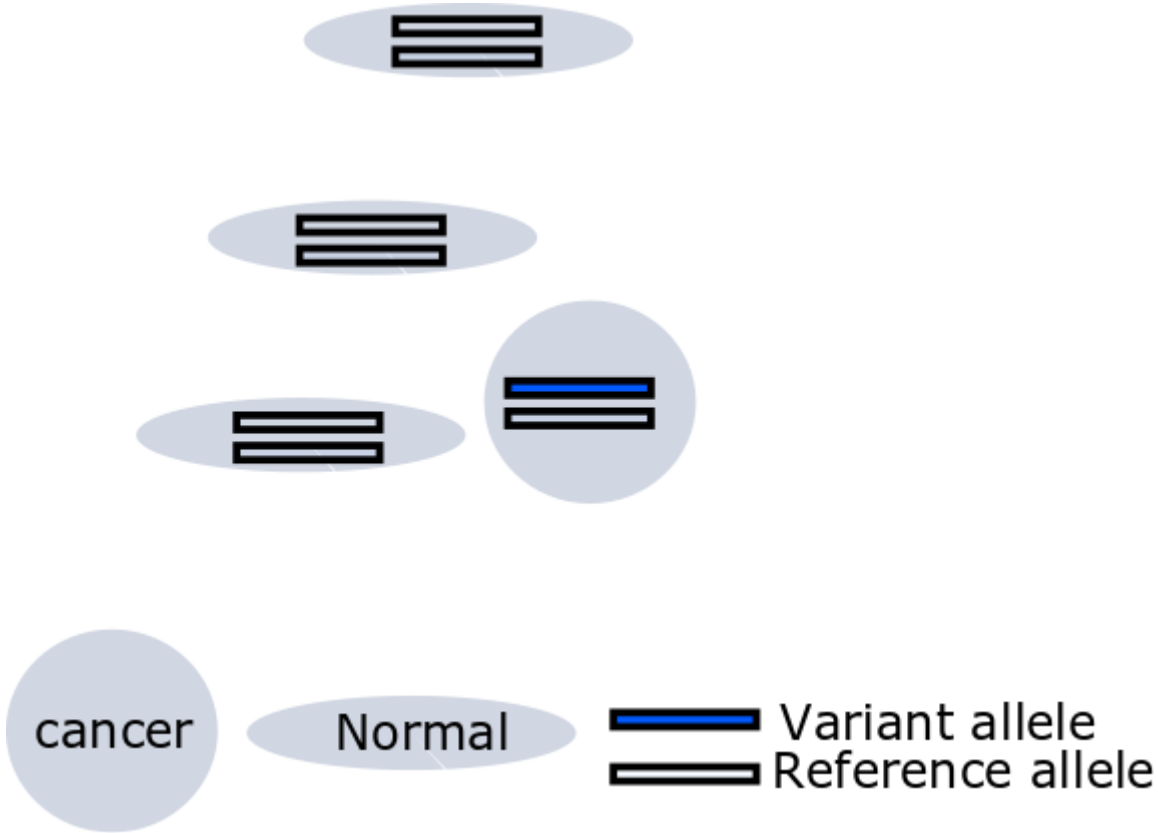
> Is it possible **64% allele frequency** of somatic mutation?

Questions

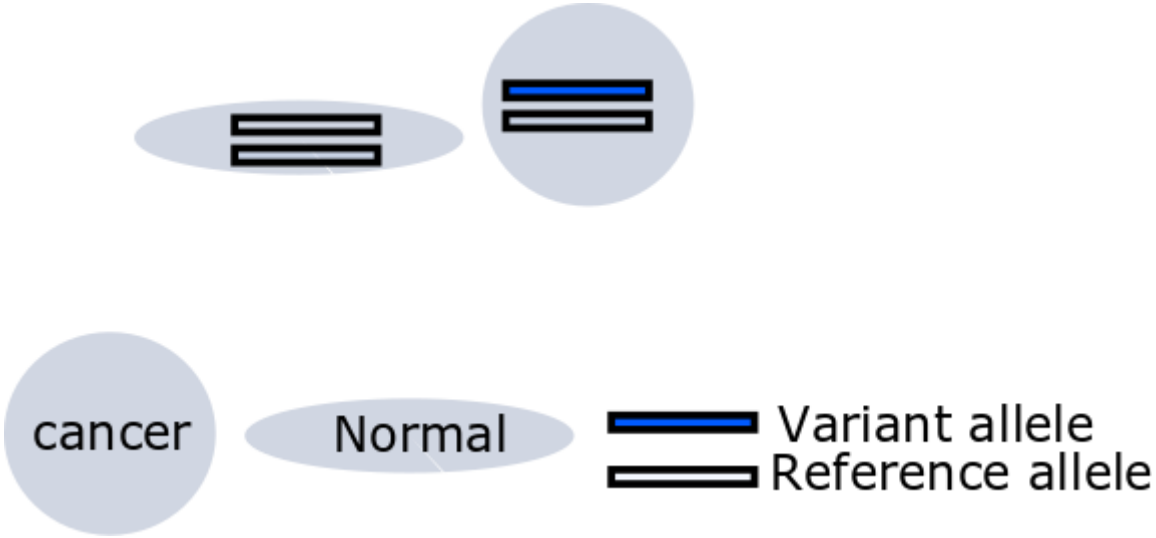
> Can we determine whether the variant is germline or somatic using allele frequency and tumor percentage in tumor only test?

Variant allele frequency in clinical tumor sample

$$Allele\ frequency \approx Read\ count\ proportion$$

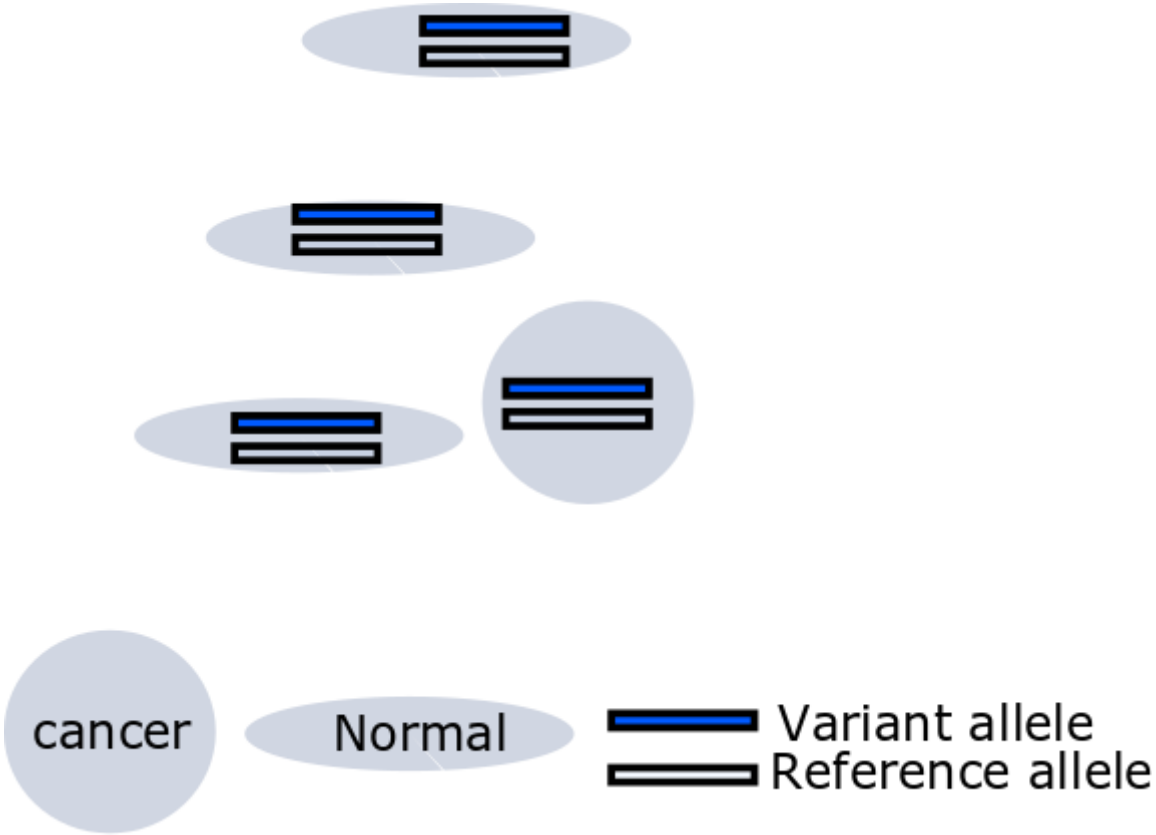

cancer
Normal
Variant allele
Reference allele

Somatic variant, Two copy, Tumor cellularity 50%



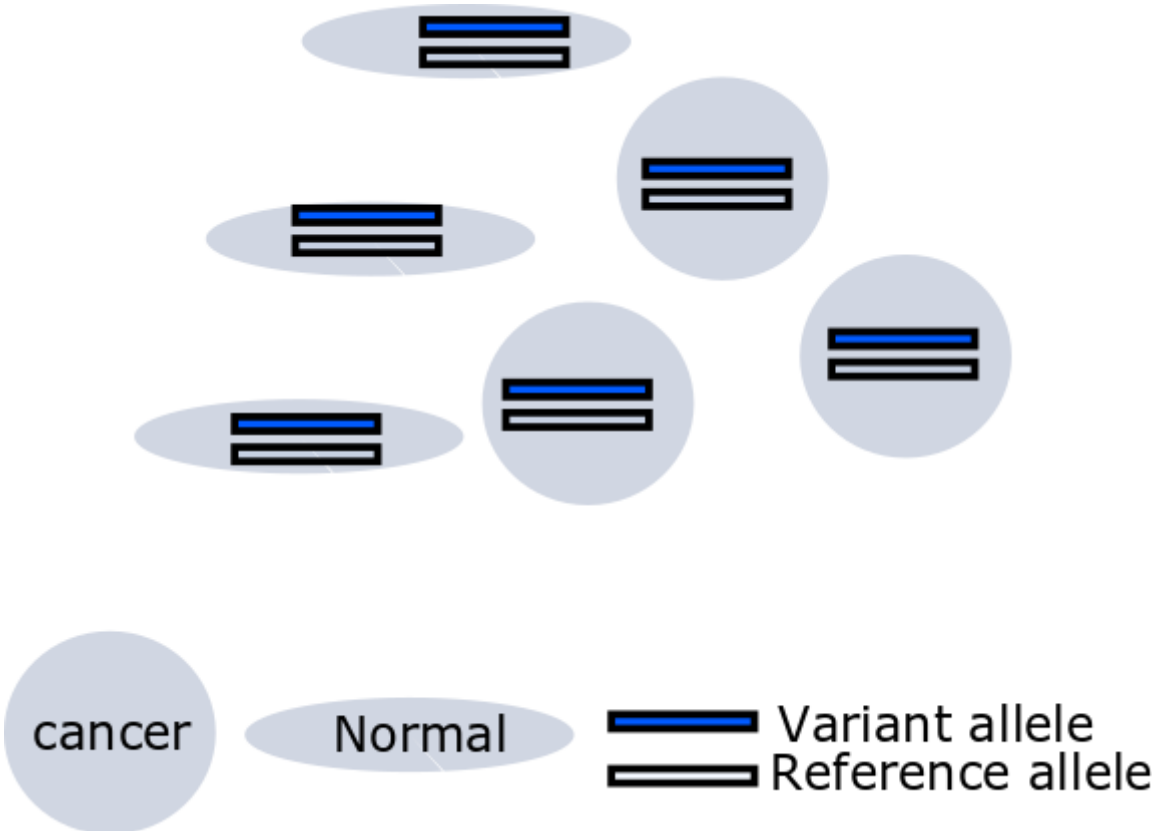
Variant allele frequency?

Germline variant, Heterozigosity, Two copy, Tumor cellularity 25%



Variant allele frequency?

Germline variant, Heterozigosity, Two copy, Tumor cellularity 50%

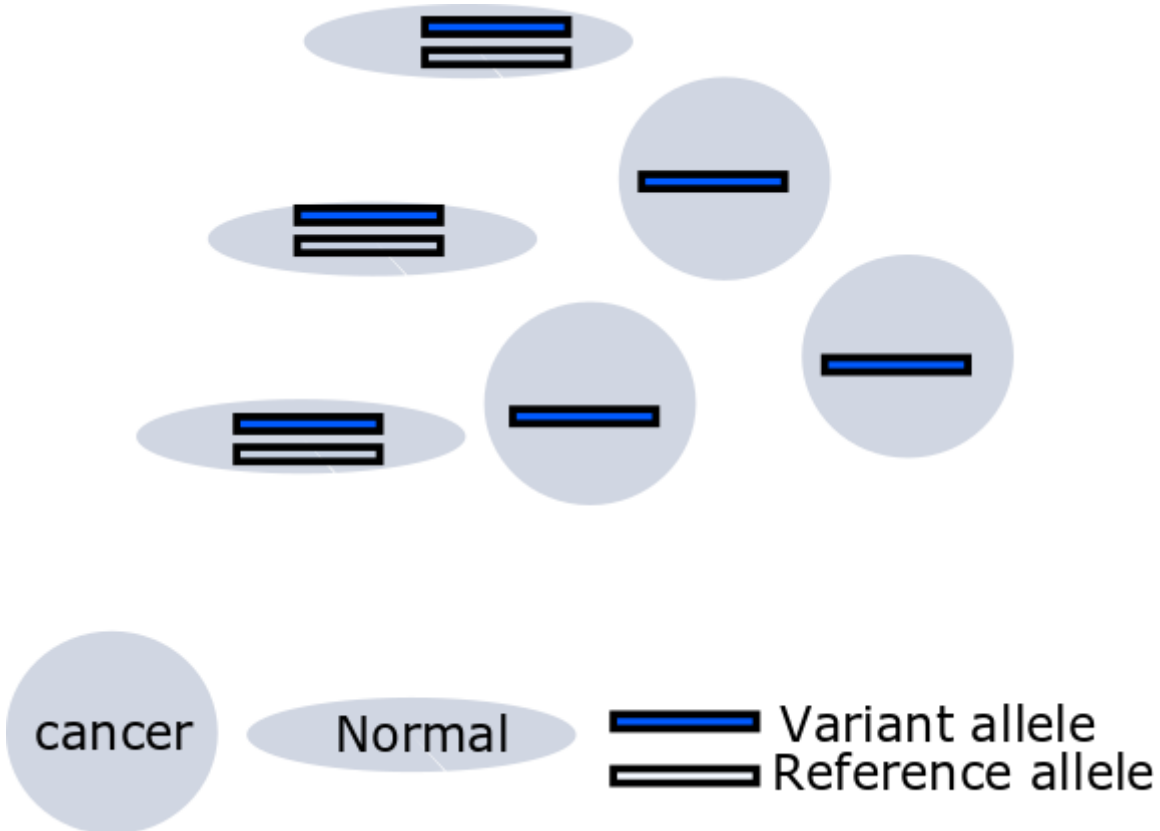


Variant allele frequency?

Proposition

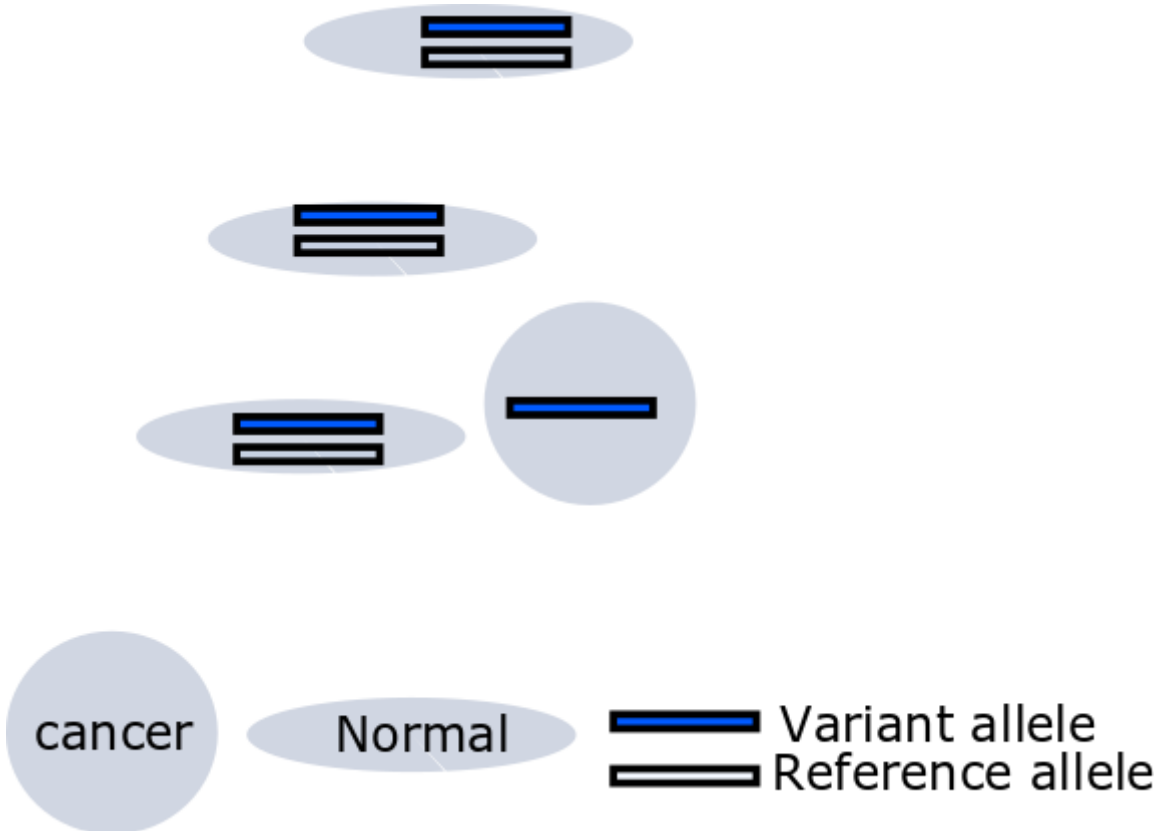
Germline, homo	Germline, hetero	Somatic
100%	50%	half of tumor cellularity, $\leq 50\%$

Germline variant, Heterozigosity, One copy, LOH, Tumor cellularity 50%



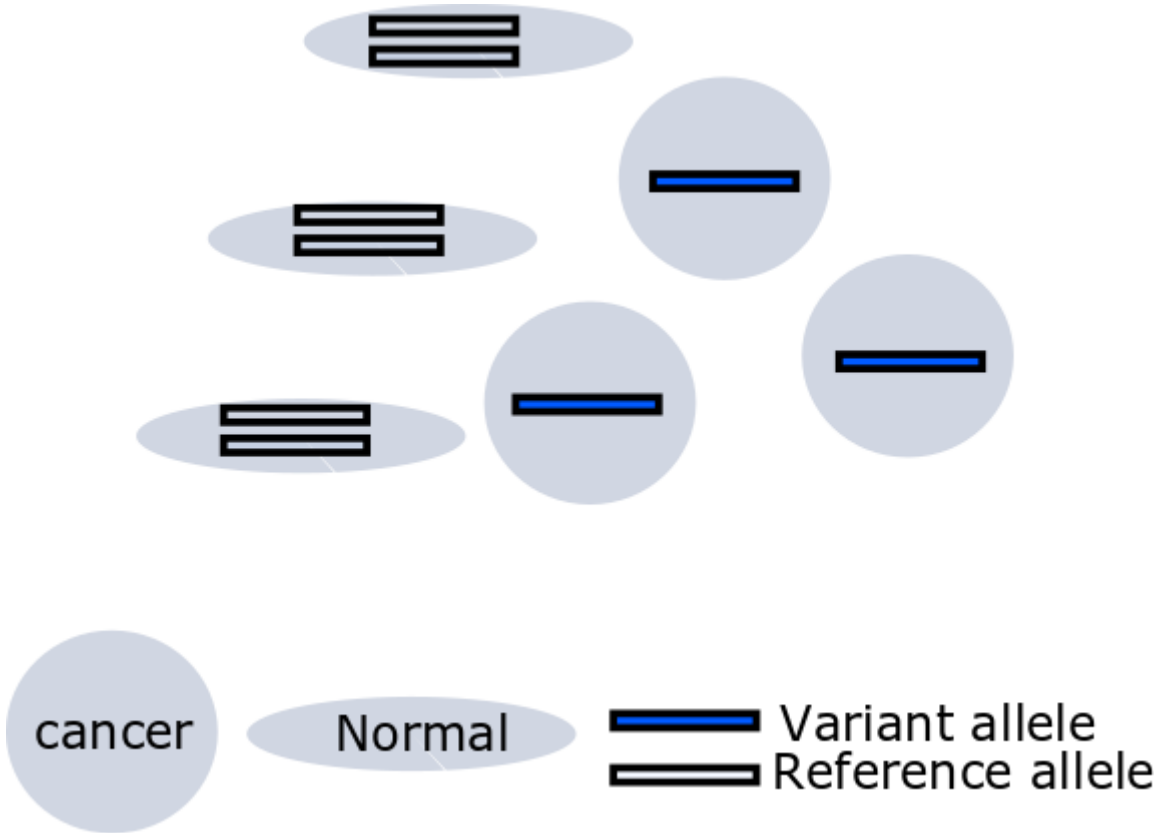
Variant allele frequency?

Germline variant, Heterozigosity, One copy, LOH, Tumor cellularity 25%



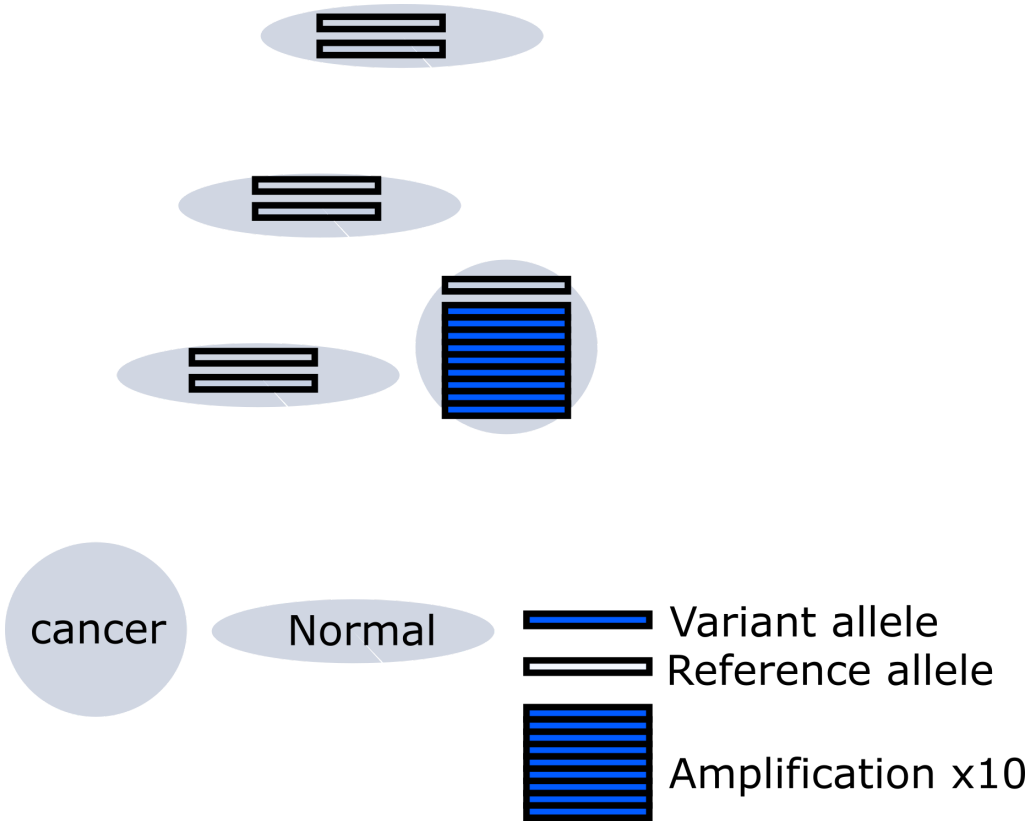
Variant allele frequency?

Somatic variant, One copy, LOH, Tumor cellularity 50%



Variant allele frequency?

Somatic variant, Amplification, Tumor cellularity 25%



Variant allele frequency?

Variant allele frequency in clinical tumor sample

- Germline vs somatic
- Tumor cell proportion
- Loss of heterozygosity
- Copy number

Allele frequency in Somatic vs Germline in tumor only sample

- BRCA1 mutation: Positive - p.Glu649Ter (c.1945G>T)
- Variant allele frequency 63.84%
- Tumor cell percentage: 70%

Allele frequency fomulas [1]

$$AF_{germline} = \frac{pV+1-p}{pC+2(1-p)}$$
$$AF_{somatic} = \frac{pV}{pC+2(1-p)}$$

- Given copy number (C)
- Variant allele count (V)
- Sample purity (p)
- Variant status (somatic or germline)

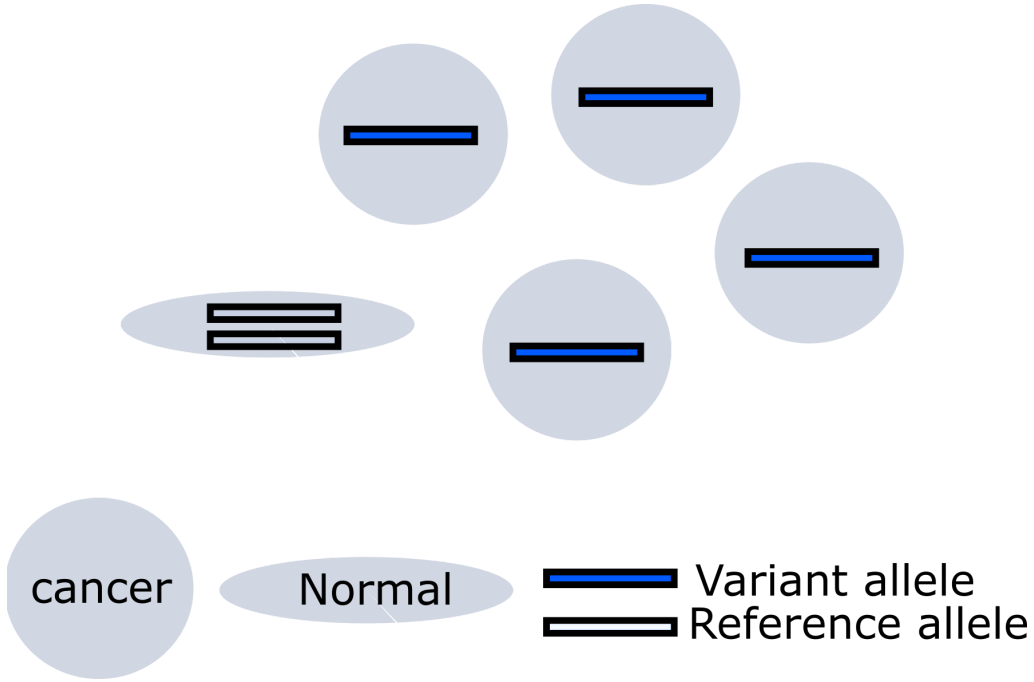
Limited information

What we know	What we don't know
Tumor cell percentage	Copy number
Variant allele frequency	LOH

Allele frequency

- BRCA1 mutation: Positive - p.Glu649Ter (c.1945G>T)
- Variant allele frequency 63.84%
- Tumor cell percentage: 70% -> 80%
- Somatic (supposed)
- LOH (supposed)
- One copy (supposed)

Allele frequency

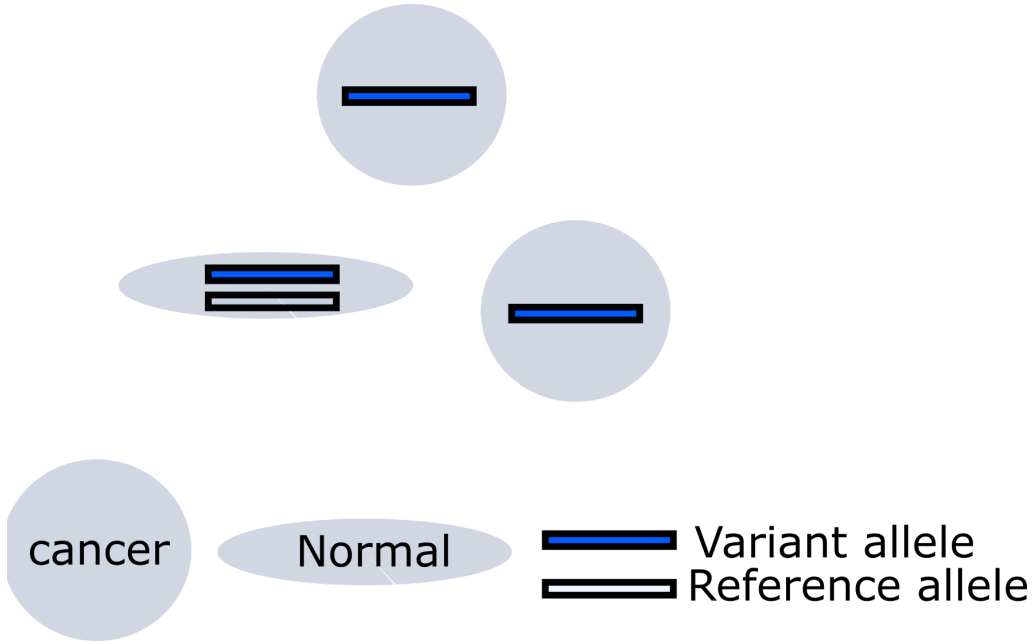


$$AF_{somatic} = \frac{pV}{pC+2(1-p)} = 0.67$$

Allele frequency

- BRCA1 mutation: Positive - p.Glu649Ter (c.1945G>T)
- Variant allele frequency 63.84%
- Tumor cell percentage: 70% -> 60%
- Germline (supposed)
- LOH (supposed)
- One copy (supposed)

Allele frequency



$$AF_{germline} = \frac{pV+1-p}{pC+2(1-p)} = 0.71$$

Caveat

- Strand bias
- Heterogeneity

Allele specific copy number

Variant pathogenecity

Question

- 76/M, Lung adenocarcinoma
- NM_000059.3(BRCA2):c.5683G>A (p.Glu1895Lys)

> This sequence change replaces glutamic acid with lysine at codon 1895 of the BRCA2 protein (p.Glu1895Lys). The glutamic acid residue is weakly conserved and there is a small physicochemical difference between glutamic acid and lysine. This variant is present in population databases (rs146351301, ExAC 0.009%). This variant has been reported in individuals affected with breast cancer (PMID: 25682074, 20104584, 27257965, 28664449). In the literature, this variant is also known as 5911G>A. ClinVar contains an entry for this variant (Variation ID: 142307). Algorithms developed to predict the effect of missense changes on protein structure and function output the following: SIFT: "Tolerated"; PolyPhen-2: "Benign"; Align-GVGD: "Class C0". The lysine amino acid residue is found in multiple mammalian species, suggesting that this missense change does not adversely affect protein function. These predictions have not been confirmed by published functional studies and their clinical significance is uncertain. In summary, the available evidence is currently insufficient to determine the role of this variant in disease. Therefore, it has been classified as a Variant of Uncertain Significance.

Sherloc Invitae

Show 8 entries

	Name	Value
1	% Frequency	33.91
2	Allele Coverage	G=1263, A=648
3	Allele Ratio	G=0.6609, A=0.3391
4	Amino Acid Change	p.Glu1895Lys
5	ClinVar	Uncertain significance
6	Coding	c.5683G>A
7	Codon	AAG
8	Coverage	1911

Showing 1 to 8 of 41 entries Previous 1 2 3 4 5 6 Next

Standards and Guidelines for the Interpretation of Sequence Variants

- To describe variants identified in Mendelian disorders
- **American College of Medical Genetics and Genomics (ACMG)**
- ENIGMA BRCA1/2 Gene Variant Classification Criteria
- International Agency for Research on Cancer (IARC)

Categories of interpretation of variants

- Pathogenic
- Likely-pathogenic
- Uncertain (VUS)
- Likely-benign
- Benign

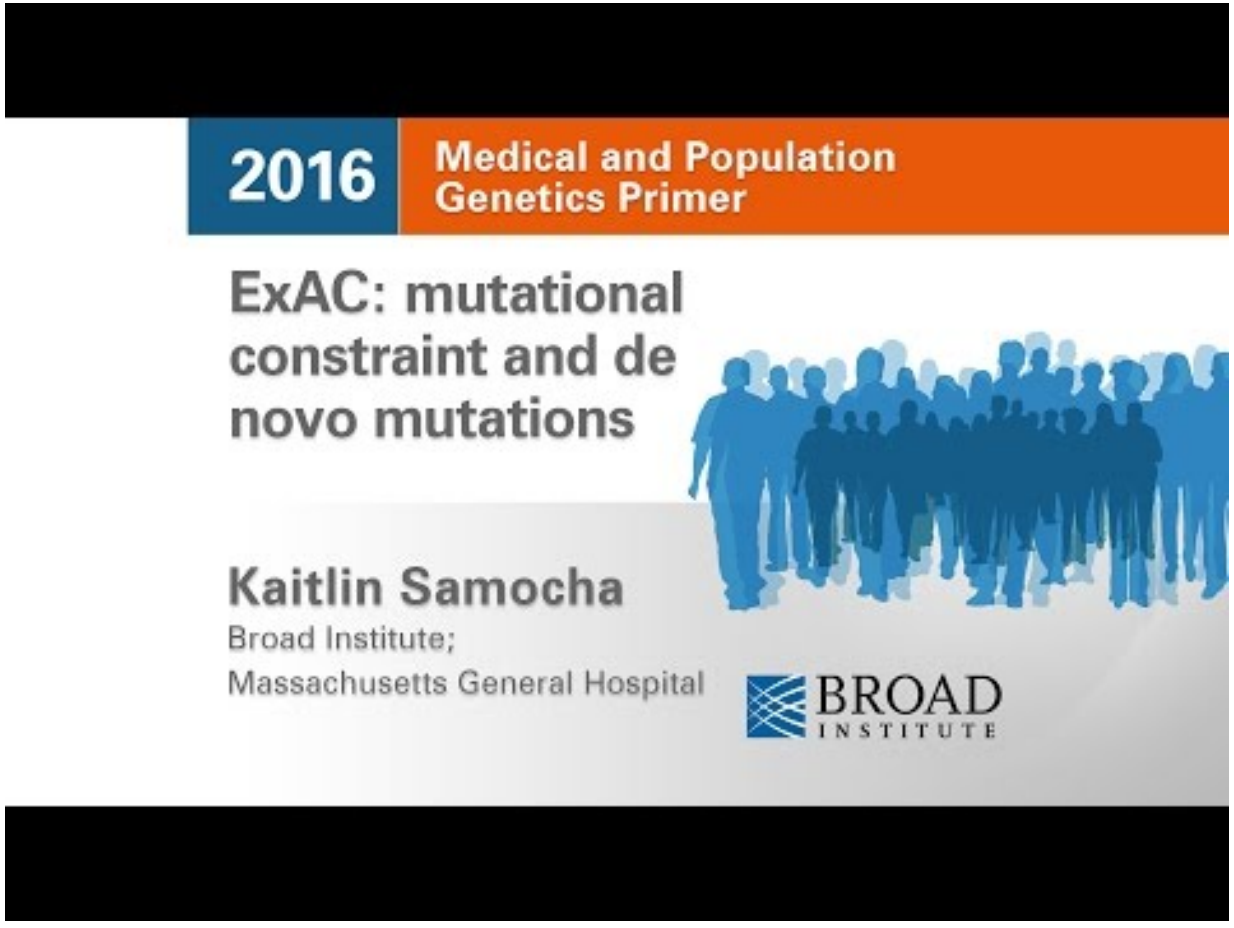
Let's guess the evidences

Famly pedigree

Segregation data (BS1, PP1)

- Caveat
  - linkage disequilibrium
  - Penetration rate
  - Difficult statistical evaluation

Population data



Population data

- 5%: benign stand alone (BA1)
- 0.5-5% (BS1)
- The first time observed variant! (Absent in population DB, PM2)

Null variant

- Frameshift, Nonsense, canonical +-1 or 2 splicing site, initiation codon
- Caveat: LOF variants at the extreme 3′ end of a gene
- Caveat: presence of multiple transcripts

Computational (in silico) data

- PolyPhen2, SIFT, MutationTaster, etc
- Mutational hot spot and/or critical and wellestablished (PM1)
- Protein length changes due to in-frame deletions/insertions and stop losses functional domain (PM4 BP3)
- Novel missense at the same position (PM5)

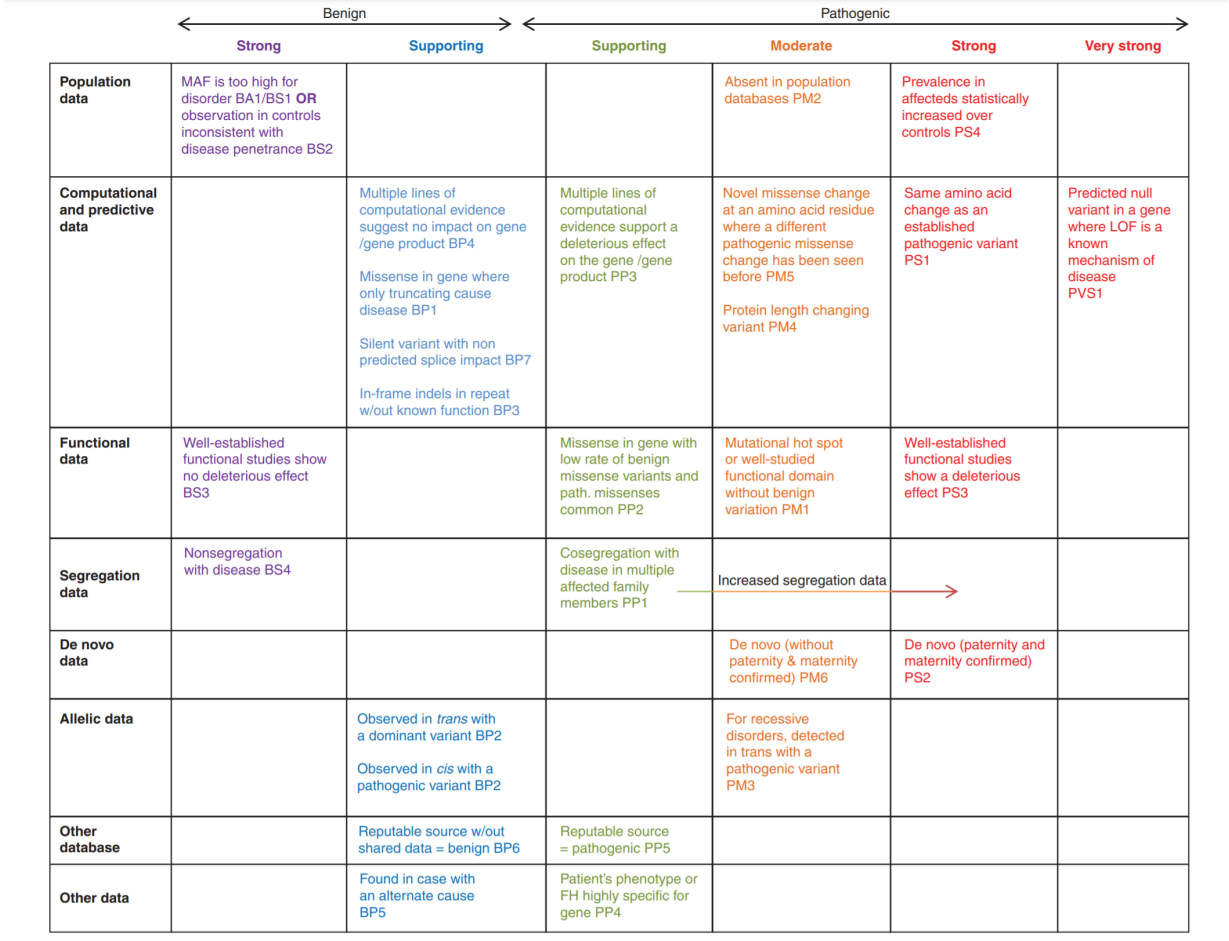
Other evidence

- de novo variants (PS2 PM6)
- Functional studies (PS3 BS3)
- Allelic data (BP2 PM3)

Evidences of interpretation

- Population data
- Computational data
- Functional data
- Segregation data
- De novo data
- Allele data
- Other databases
- Other data

27 variant attributes



	Benign		Pathogenic			
	Strong	**Supporting**	**Supporting**	**Moderate**	**Strong**	**Very strong**
Population data	MAF is too high for disorder BA1/BS1 **OR** observation in controls inconsistent with disease penetrance BS2			Absent in population databases PM2	Prevalence in affecteds statistically increased over controls PS4	
Computational and predictive data		Multiple lines of computational evidence suggest no impact on gene /gene product BP4 Missense in gene where only truncating cause disease BP1 Silent variant with non predicted splice impact BP7 In-frame indels in repeat w/out known function BP3	Multiple lines of computational evidence support a deleterious effect on the gene /gene product PP3	Novel missense change at an amino acid residue where a different pathogenic missense change has been seen before PM5 Protein length changing variant PM4	Same amino acid change as an established pathogenic variant PS1	Predicted null variant in a gene where LOF is a known mechanism of disease PVS1
Functional data	Well-established functional studies show no deleterious effect BS3		Missense in gene with low rate of benign missense variants and path. missenses common PP2	Mutational hot spot or well-studied functional domain without benign variation PM1	Well-established functional studies show a deleterious effect PS3	
Segregation data	Nonsegregation with disease BS4		Cosegregation with disease in multiple affected family members PP1	Increased segregation data →		
De novo data				De novo (without paternity & maternity confirmed) PM6	De novo (paternity and maternity confirmed) PS2	
Allelic data		Observed in *trans* with a dominant variant BP2 Observed in *cis* with a pathogenic variant BP2		For recessive disorders, detected in trans with a pathogenic variant PM3		
Other database		Reputable source w/out shared data = benign BP6	Reputable source = pathogenic PP5			
Other data		Found in case with an alternate cause BP5	Patient's phenotype or FH highly specific for gene PP4			

Figure 1 Evidence framework. This chart organizes each of the criteria by the type of evidence as well as the strength of the criteria for a benign (left side) or pathogenic (right side) assertion. Evidence code descriptions can be found in **Tables 3** and **4**. BS, benign strong; BP, benign supporting; FH, family history; LOF, loss of function; MAF, minor allele frequency; path., pathogenic; PM, pathogenic moderate; PP, pathogenic supporting; PS, pathogenic strong; PVS, pathogenic very strong.

Characteristics of BRCA1/2

- LOF known mechanism of disease (for PVS1)
- Mode of inheritance (for PM3/BP2)
  - AD/AR (BRCA2)
- Missense pathogenic (for PP2/BP1)
  - BRCA2 1%
- Hot spot or critical/well-established functional domain (for PM1)
  - BRCA2, Helical (2479-2667), OB (2670-2799 and 3052-3190), Tower (2831-2872)

NM_000059.3(BRCA2):c.5683G>A (p.Glu1895Lys)

- LOF known mechanism of disease (for PVS1)
- Mode of inheritance (for PM3/BP2)
  - AD/AR (BRCA2)
- Missense pathogenic (for PP2/**BP1**)
  - **BRCA2 1%**
- Hot spot or critical/well-established functional domain (for PM1)
  - BRCA2, Helical (2479-2667), OB (2670-2799 and 3052-3190), Tower (2831-2872)

Characteristics of varant (NM_000059.3(BRCA2):c.5683G>A (p.Glu1895Lys))

- ClinVar (Uncertain significance (Last evaluated: Nov 1, 2015)) (PP5, BP6)
- Population AF: $8.29e^{-06}$ (PM2, BA1, BS1)
- Insilico SIFT 1.0, phyloP -0.72, PolyPhen-2 0.004 **BP4**

Functional assay (BS3)

- Findlay (2018) Nature 562: 217 PubMed: 30209399
- Guidugli (2013) Cancer Res 73: 265 PubMed: 23108138
- Biswas (2011) Blood 118: 2430 PubMed: 21719596
- Becker (2012) Breast Cancer Res Treat 135: 167 PubMed: 22729890

> This sequence change replaces glutamic acid with lysine at codon 1895 of the BRCA2 protein (p.Glu1895Lys). The glutamic acid residue is weakly conserved and there is a small physicochemical difference between glutamic acid and lysine. This variant is present in population databases (rs146351301, ExAC 0.009%). This variant has been reported in individuals affected with breast cancer (PMID: 25682074, 20104584, 27257965, 28664449). In the literature, this variant is also known as 5911G>A. ClinVar contains an entry for this variant (Variation ID: 142307). Algorithms developed to predict the effect of missense changes on protein structure and function output the following: SIFT: "Tolerated"; PolyPhen-2: "Benign"; Align-GVGD: "Class C0". The lysine amino acid residue is found in multiple mammalian species, suggesting that this missense change does not adversely affect protein function. These predictions have not been confirmed by published functional studies and their clinical significance is uncertain. In summary, the available evidence is currently insufficient to determine the role of this variant in disease. Therefore, it has been classified as a Variant of Uncertain Significance.

Sherloc Invitae

BRCA2 c.10150C>T p.Arf3384Ter

KOBRA [2]

End truncation

Clinvar

BRCA2 c.9976A>T (p.Lys3326*) variant, located upstream of this variant and also in the last exon of the gene, is a known benign variant.

Conclusions

- Allele frequency
- ACMG guideline

References

[1] J. X. Sun, Y. He, E. Sanford, et al. "A computational approach to distinguish somatic vs. germline origin of genomic alterations from deep sequencing of cancer specimens without a matched normal". En. In: *PLOS Computational Biology* 14.2 (2018), p. e1005965. ISSN: 1553-7358.

[2] E. Kang, M. Seong, S. K. Park, et al. "The prevalence and spectrum of BRCA1 and BRCA2 mutations in Korean population: recent update of the Korean Hereditary Breast Cancer (KOHBRA) study". En. In: *Breast Cancer Res Treat* 151.1 (5. 2015), pp. 157-168. ISSN: 1573-7217.