Practical - Variant Interpretation in a Clinical Laboratory

Best Practice

Natalie Groves
Final Year Trainee Bioinformatician
Sheffield Diagnostic Genetics Service

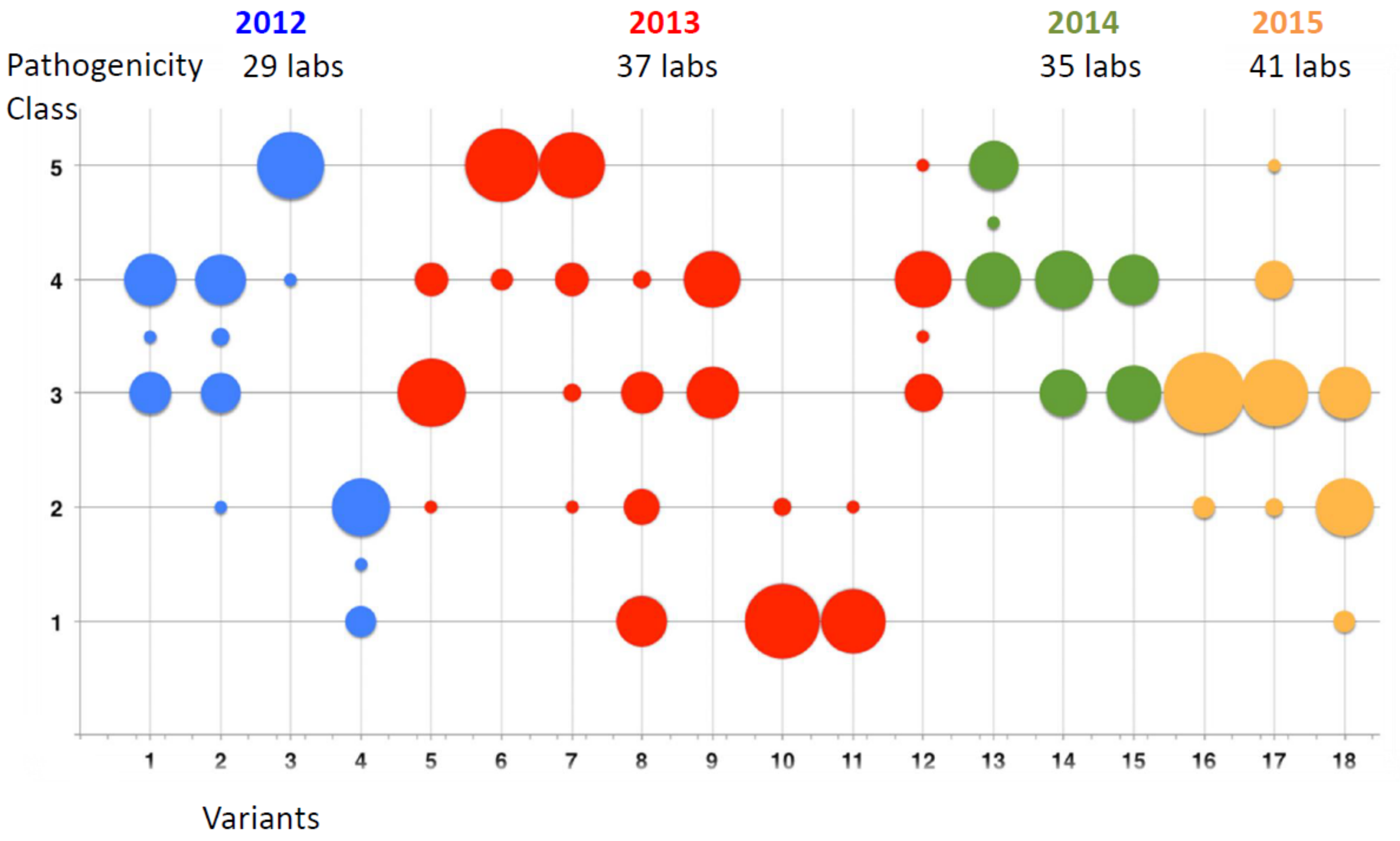
Sheffield Children's NHS
NHS Foundation Trust

Variant Interpretation

- You've done the test and got a result but what does it mean..?
  - Inheritance pattern?
  - Phenotype?
  - Pathogenic effects?
- Best practice
  - To standardise assessment of variants across clinical laboratories (in theory)
  - In practice classification still varies among laboratories

Consistent inconsistency?

Sheffield Children's NHS
NHS Foundation Trust



Best Practice Guidelines

- Association for Clinical Genetic Science 2015
  - Defines the pieces of information that should be collected
  - Gives some guidance for each classification

- American College of Medical Genetics and Genomics 2015
  - Designed to reduce the number of "causative" predictions that does not have sufficient supporting evidence
  - Recommended for adoption in UK Nov 2016

What are we looking at?

- HGVS nomenclature
  - Notation is dependent on type of variation and the level of description (e.g. genomic, coding DNA, protein)
  - Must detail the reference sequence used
    - g. = genomic
    - m. = mitochondrial
    - c. = coding DNA
    - n. = non-coding DNA
    - p. = protein
  - Must give the position(s)
  - Must give the reference and alternative nucleotide/amino acid

g.45576A>C

"A" nucleotide at genomic coordinate 45576 has been substituted for a "C"

c.56G>T

"G" nucleotide at cDNA coordinate 56 has been substituted for a "T"

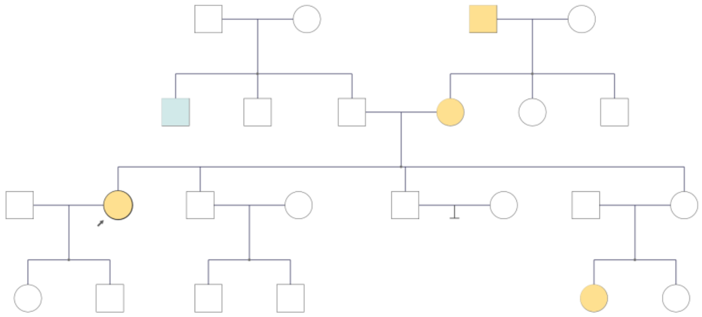
p.(Trp26Cys)

Tryptophan residue at position 26 is changed to a Cysteine *brackets indicate predicted consequence

Who are we looking at?

Sheffield Children's NHS
NHS Foundation Trust

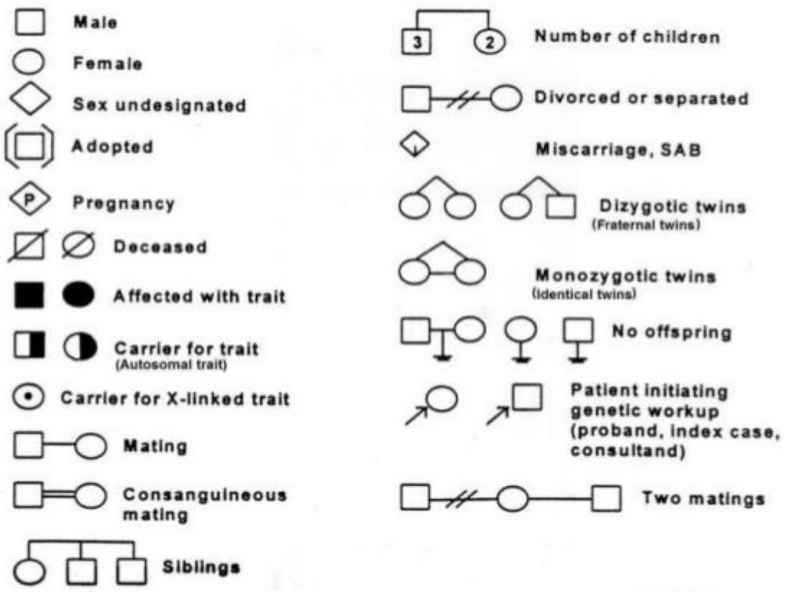
- Test requested will narrow the number of genes being assessed (for a targeted panel)
- Still identify a number of variants
- Phenotype and family information can help to triage the variant list



A pedigree diagram can help determine inheritance – particularly if other family members have had testing or show clinical features

Drawing a pedigree

Sheffield Children's NHS
NHS Foundation Trust



Symbols used when drawing a genetic pedigree

- Describe the family structure and any affected individuals
- Usually includes proband's grandparents
- Affected family members must have full clinical details (DOB, condition, relevant testing, known mutations)

https://playground.phenotips.org/Families/Create

Your turn…

- Fill out the phenotype section of the form using the information provided in the patient sheet
- Try to identify HPO terms for the descriptions given
- Generate a pedigree from the description of the family on phenotips
- Try to identify a likely inheritance pattern from the pedigree

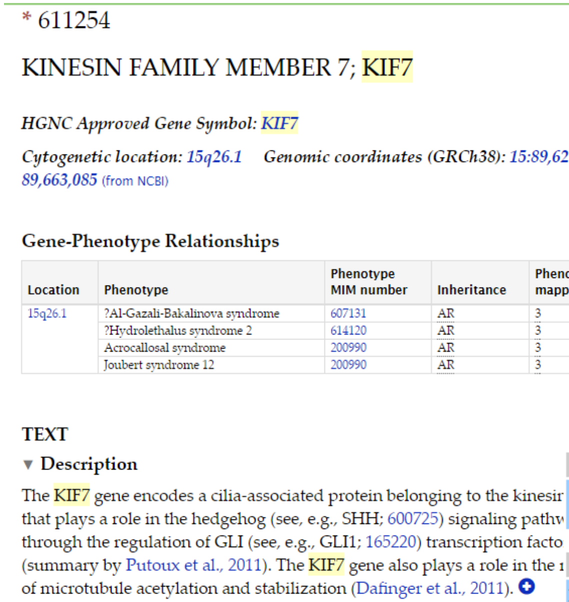
What would pathogenicity mean?

- It is important to understand what would be expected if a mutation was found in the gene you are investigating
  - What disorder(s)?
  - What phenotype?
  - What inheritance pattern?
  - What type of mutations?
- If the information gathered about the patient does not match it can suggest a polymorphic role

Useful Resources

Sheffield Children's NHS
NHS Foundation Trust

OMIM



* 611254

KINESIN FAMILY MEMBER 7; KIF7

HGNC Approved Gene Symbol: KIF7

Cytogenetic location: 15q26.1 *Genomic coordinates (GRCh38): 15:89,627,969-89,663,085* (from NCBI)

Gene-Phenotype Relationships

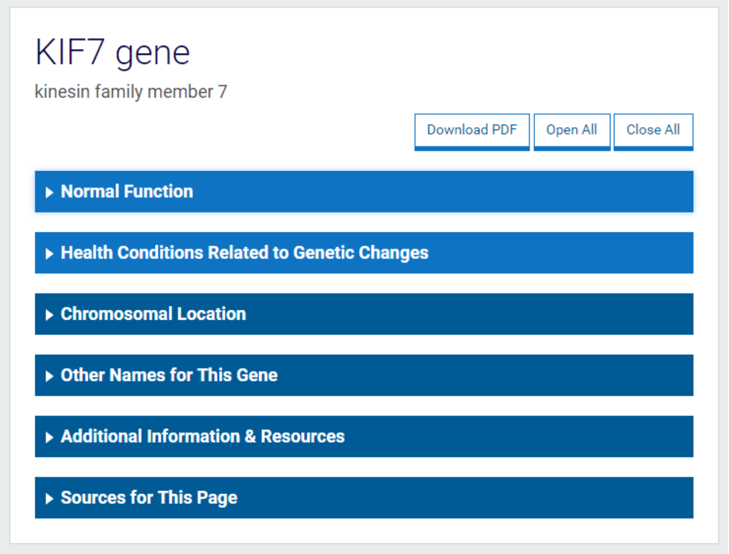
Location	Phenotype	Phenotype MIM number	Inheritance	Phenotype mapping key
15q26.1	?Al-Gazali-Bakalinova syndrome	607131	AR	3
	?Hydrolethalus syndrome 2	614120	AR	3
	Acrocallosal syndrome	200990	AR	3
	Joubert syndrome 12	200990	AR	3

TEXT

▼ Description

The KIF7 gene encodes a cilia-associated protein belonging to the kinesir that plays a role in the hedgehog (see, e.g., SHH; 600725) signaling pathw through the regulation of GLI (see, e.g., GLI1; 165220) transcription facto (summary by Putoux et al., 2011). The KIF7 gene also plays a role in the r of microtubule acetylation and stabilization (Dafinger et al., 2011).

Genetics Home Reference

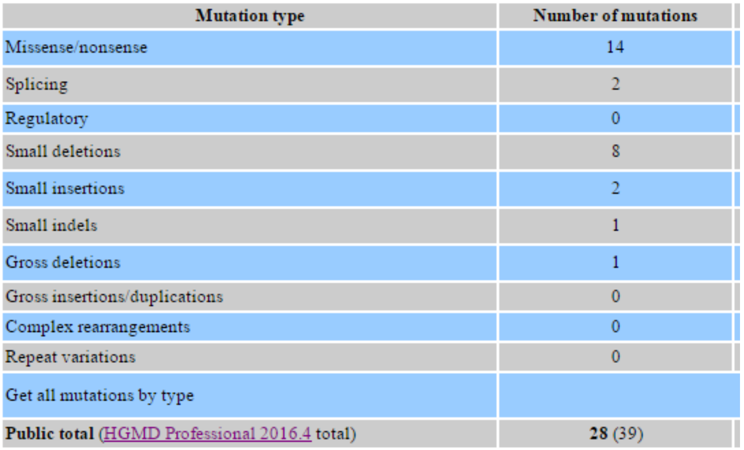


KIF7 gene

kinesin family member 7

Download PDF
Open All
Close All
▸ Normal Function
▸ Health Conditions Related to Genetic Changes
▸ Chromosomal Location
▸ Other Names for This Gene
▸ Additional Information & Resources
▸ Sources for This Page

Gene Symbol	Chromosomal location	Gene name
KIF7 (Aliases: available to subscribers)	15q26.1	Kinesin family membe (Aliases: available to subscribe



Mutation type	Number of mutations
Missense/nonsense	14
Splicing	2
Regulatory	0
Small deletions	8
Small insertions	2
Small indels	1
Gross deletions	1
Gross insertions/duplications	0
Complex rearrangements	0
Repeat variations	0
Get all mutations by type	
Public total (HGMD Professional 2016.4 total)	**28** (39)

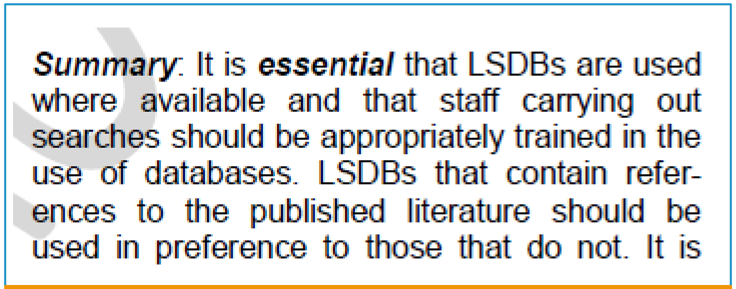
HGMD

Your turn…

- Fill out the disease background section of the form for the gene given on your patient information sheet
- Try to identify the inheritance pattern, phenotype associated with the disorder and whether there are certain types of mutation that are more common

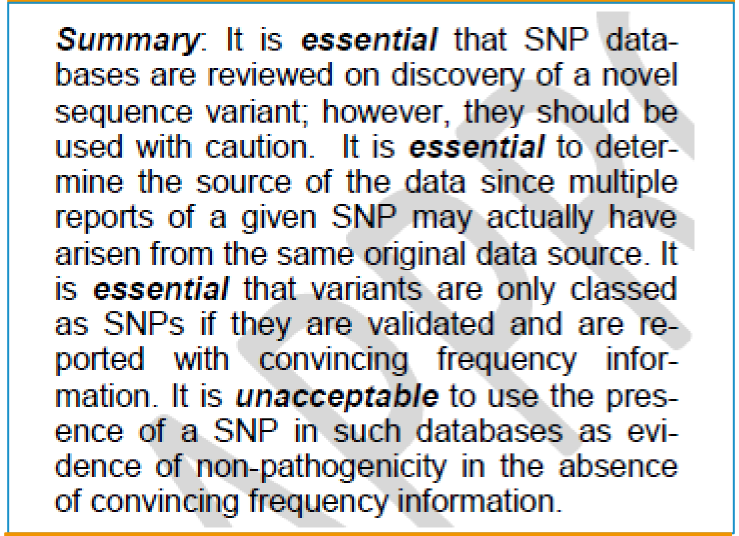
Has anyone seen it before?

Variant Databases



Summary: It is ***essential*** that LSDBs are used where available and that staff carrying out searches should be appropriately trained in the use of databases. LSDBs that contain references to the published literature should be used in preference to those that do not. It is

4.1 Variant databases including LSDBs



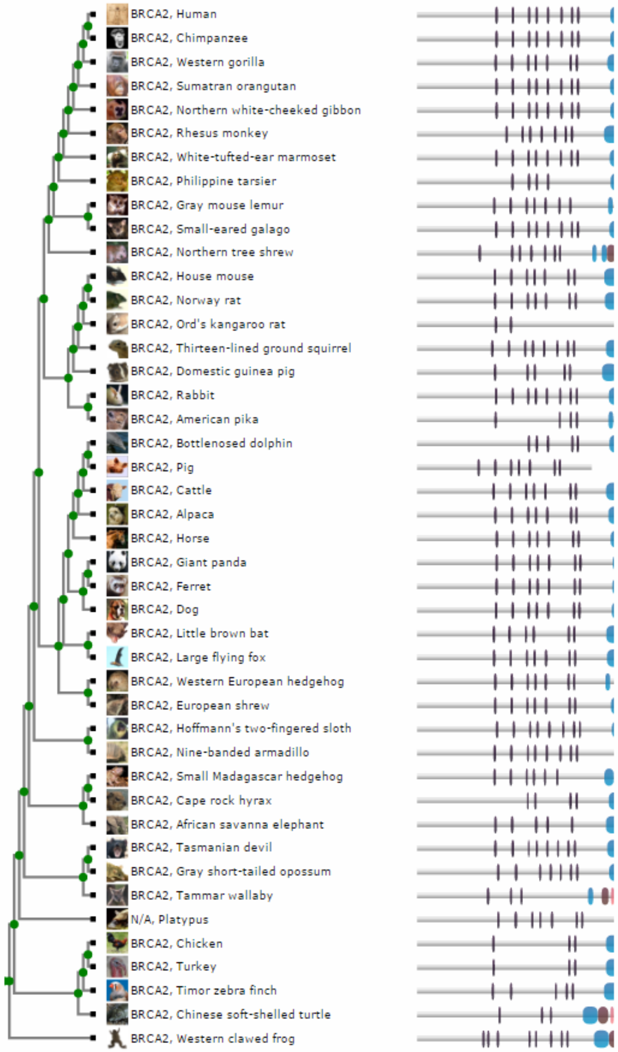
Summary: It is ***essential*** that SNP databases are reviewed on discovery of a novel sequence variant; however, they should be used with caution. It is ***essential*** to determine the source of the data since multiple reports of a given SNP may actually have arisen from the same original data source. It is ***essential*** that variants are only classed as SNPs if they are validated and are reported with convincing frequency information. It is ***unacceptable*** to use the presence of a SNP in such databases as evidence of non-pathogenicity in the absence of convincing frequency information.

4.2 Presence or absence on SNP databases

- Can be used to determine whether a variant is likely to be pathogenic
- Now contain pathogenic variants
- Some variants added with little supporting evidence
  - Artefact of sequencing
  - Insufficient evidence to determine role

Is the position important?

- Variants in functional domains or involved in protein structure are more likely to be pathogenic
- Sequence conservation can indicate sites that cannot tolerate variation


BRCA2, Human
BRCA2, Chimpanzee
BRCA2, Western gorilla
BRCA2, Sumatran orangutan
BRCA2, Northern white-cheeked gibbon
BRCA2, Rhesus monkey
BRCA2, White-tufted-ear marmoset
BRCA2, Philippine tarsier
BRCA2, Gray mouse lemur
BRCA2, Small-eared galago
BRCA2, Northern tree shrew
BRCA2, House mouse
BRCA2, Norway rat
BRCA2, Ord's kangaroo rat
BRCA2, Thirteen-lined ground squirrel
BRCA2, Domestic guinea pig
BRCA2, Rabbit
BRCA2, American pika
BRCA2, Bottlenosed dolphin
BRCA2, Pig
BRCA2, Cattle
BRCA2, Alpaca
BRCA2, Horse
BRCA2, Giant panda
BRCA2, Ferret
BRCA2, Dog
BRCA2, Little brown bat
BRCA2, Large flying fox
BRCA2, Western European hedgehog
BRCA2, European shrew
BRCA2, Hoffmann's two-fingered sloth
BRCA2, Nine-banded armadillo
BRCA2, Small Madagascar hedgehog
BRCA2, Cape rock hyrax
BRCA2, African savanna elephant
BRCA2, Tasmanian devil
BRCA2, Gray short-tailed opossum
BRCA2, Tammar wallaby
N/A, Platypus
BRCA2, Chicken
BRCA2, Turkey
BRCA2, Timor zebra finch
BRCA2, Chinese soft-shelled turtle
BRCA2, Western clawed frog

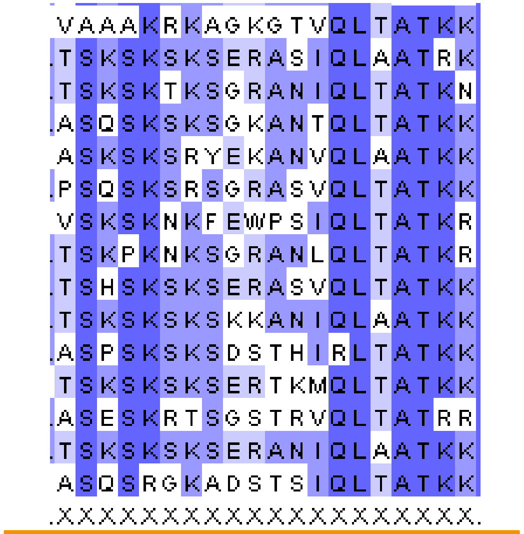
TreeFam gene tree for BRCA2
http://www.treefam.org/family/TF105041#tabview=tab1

Your turn…

- Look at the gene in pfam
- What domains does it have?
- Is your variant in a domain that is important for function or structure?

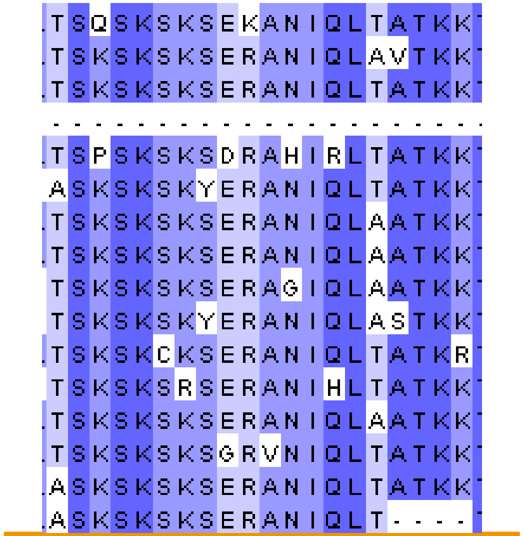
Multiple Sequence Alignment

- Extremely sensitive to diversity and gaps
- Divergence of the region directly surrounding the variant is important
- Quality of MSA is assessed based on the probability of the residue showing a high conservation due to chance
- Different amino acids don't always indicate tolerance
- Similarity doesn't always indicate pathogenicity



```
VAAAKRKAGKGTVQLTATKK
TSKSKSKSERASIQLAATRK
TSKSKTKSGRANIQLTATKN
ASQSKSKSGKANTQLTATKK
ASKSKSRYEKANVQLAATKK
PSQSKSRSGRASVQLTATKK
VSKSKNKFEWPSIQLTATKR
TSKPKNKSGRANLQLTATKR
TSHSKSKSERASVQLTATKK
TSKSKSKSKKANIQLAATKK
ASPSKSKSDSTHIRLTATKK
TSKSKSKSERTKMQLTATKK
ASESKRTSGSTRVQLTATRR
TSKSKSKSERANIQLAATKK
ASQSRGKADSTSIQLTATKK
.XXXXXXXXXXXXXXXXXXXX.
```

Section of variation

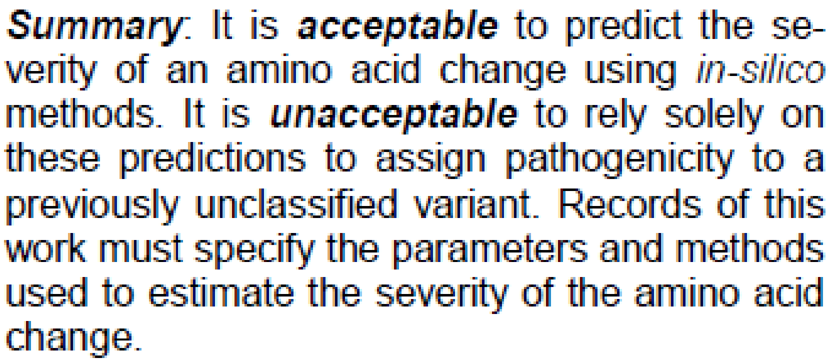


```
TSQSKSKSEKANIQLTATKK
TSKSKSKSERANIQLAVTKK
TSKSKSKSERANIQLTATKK
- - - - - - - - - - - - - - - - - - - -
TSPSKSKSDRAHIRLTATKK
ASKSKSKYERANIQLTATKK
TSKSKSKSERANIQLAATKK
TSKSKSKSERANIQLAATKK
TSKSKSKSERAGIQLAATKK
TSKSKSKYERANIQLASTKK
TSKSKCKSERANIQLTATKR
TSKSKSRSERANIHLTATKK
TSKSKSKSERANIQLAATKK
TSKSKSKSGRVNIQLTATKK
ASKSKSKSERANIQLTATKK
ASKSKSKSERANIQLT----
```

Section of similarity

Can we predict the effect?

- Computational tools have been designed to predict the impact of missense changes
- Studies have shown that the optimal tool depends on the gene and recommend a combination of tools for analysis
- Many rely on multiple sequence alignments
  - Poor alignments produce poor predictions - usually false negatives

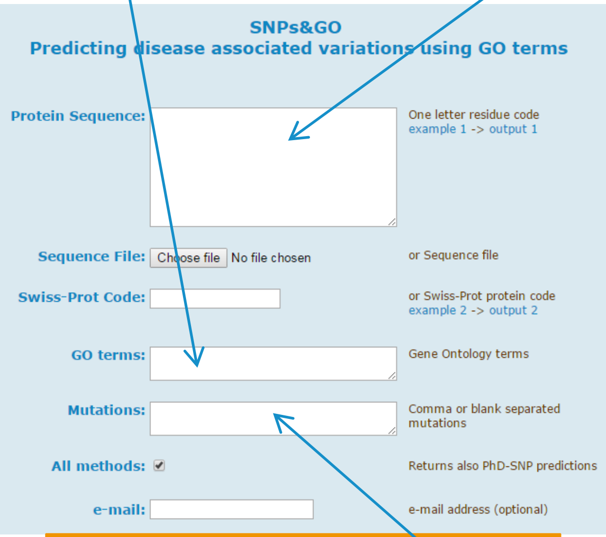


Summary: It is ***acceptable*** to predict the severity of an amino acid change using *in-silico* methods. It is ***unacceptable*** to rely solely on these predictions to assign pathogenicity to a previously unclassified variant. Records of this work must specify the parameters and methods used to estimate the severity of the amino acid change.

4.8 *In silico* prediction of pathogenic effect

Can we predict the effect?

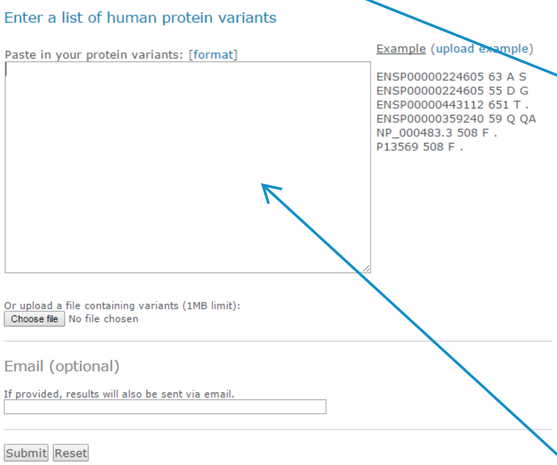
Sheffield Children's NHS
NHS Foundation Trust



SNPs&GO
Predicting disease associated variations using GO terms

Protein Sequence:		One letter residue code example 1 -> output 1
Sequence File:	Choose file No file chosen	or Sequence file
Swiss-Prot Code:		or Swiss-Prot protein code example 2 -> output 2
GO terms:		Gene Ontology terms
Mutations:		Comma or blank separated mutations
All methods:	☑	Returns also PhD-SNP predictions
e-mail:		e-mail address (optional)

SNPs&GO



Enter a list of human protein variants

Paste in your protein variants: [format]

Example (upload example)

```
ENSP00000224605 63 A S
ENSP00000224605 55 D G
ENSP00000443112 651 T .
ENSP00000359240 59 Q QA
NP_000483.3 508 F .
P13569 508 F .
```

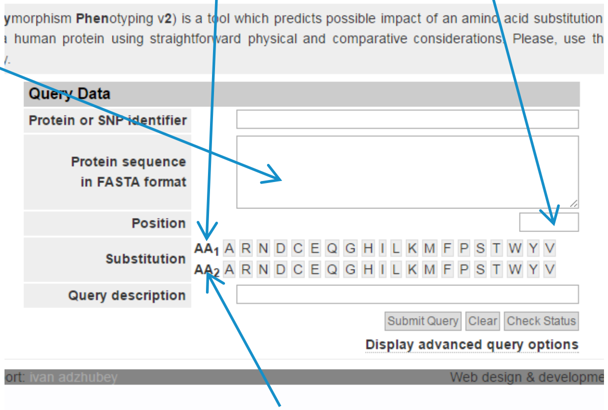
Or upload a file containing variants (1MB limit):
Choose file No file chosen

Email (optional)

If provided, results will also be sent via email.

Submit Reset

PROVEAN



ymorphism **Phen**otyping v**2**) is a tool which predicts possible impact of an amino acid substitution
ı human protein using straightforward physical and comparative considerations. Please, use th
/.

Query Data

Protein or SNP identifier	
Protein sequence in FASTA format	
Position	
Substitution	AA_1 A R N D C E Q G H I L K M F P S T W Y V AA_2 A R N D C E Q G H I L K M F P S T W Y V
Query description	

Submit Query Clear Check Status

Display advanced query options

ort: ivan adzhubey Web design & developme

PolyPhen2

Fasta sequence

GO terms

Reference amino acid

Position

Alternate amino acid

Mutation in the format (XposY)

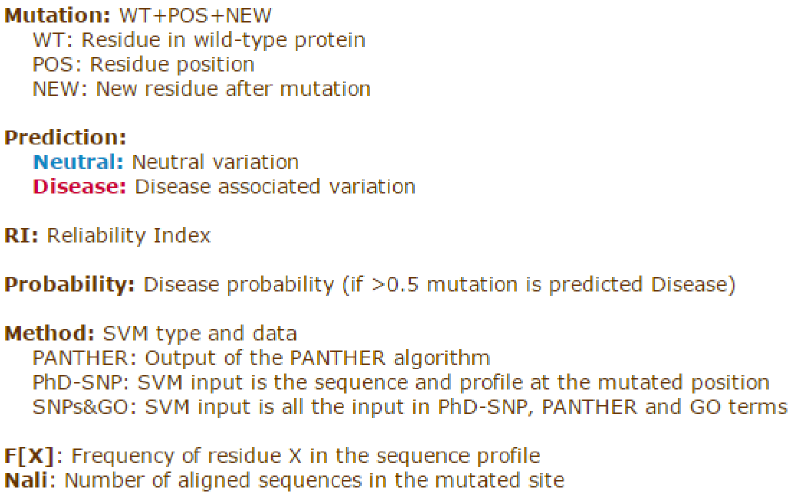
Variant in format
<protein_acc> <pos> <ref> <alt>

Your turn…

- Get the protein sequence in fasta format from NCBI (the link is on your form) using the accession given in the patient information
- Submit your variants to the online tools
- Investigate some of the GO terms associated with your gene

SNPs&GO

- Uses:
  - The effect of substitution (including nearby residues)
  - Multiple sequence alignment
  - Functional gene ontology (GO terms)
  - PANTHER prediction (evolutionary conservation)
- Reliability index shows how likelihood of accurate prediction



Mutation: WT+POS+NEW
WT: Residue in wild-type protein
POS: Residue position
NEW: New residue after mutation

Prediction:
Neutral: Neutral variation
Disease: Disease associated variation

RI: Reliability Index

Probability: Disease probability (if >0.5 mutation is predicted Disease)

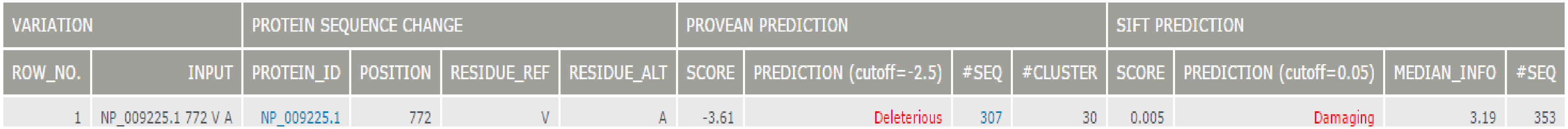
Method: SVM type and data
PANTHER: Output of the PANTHER algorithm
PhD-SNP: SVM input is the sequence and profile at the mutated position
SNPs&GO: SVM input is all the input in PhD-SNP, PANTHER and GO terms

F[X]: Frequency of residue X in the sequence profile
Nali: Number of aligned sequences in the mutated site

Output fields from SNPs&GO

PROVEAN

- Uses multiple sequence alignment to assess level of conservation at that position
- SIFT is predecessor of PROVEAN
- SIFT median info score indicates alignment quality (above 3.25 suggests there is not enough diversity for accurate prediction)



VARIATION		PROTEIN SEQUENCE CHANGE				PROVEAN PREDICTION				SIFT PREDICTION			
ROW_NO.	INPUT	PROTEIN_ID	POSITION	RESIDUE_REF	RESIDUE_ALT	SCORE	PREDICTION (cutoff=-2.5)	#SEQ	#CLUSTER	SCORE	PREDICTION (cutoff=0.05)	MEDIAN_INFO	#SEQ
1	NP_009225.1 772 V A	NP_009225.1	772	V	A	-3.61	Deleterious	307	30	0.005	Damaging	3.19	353

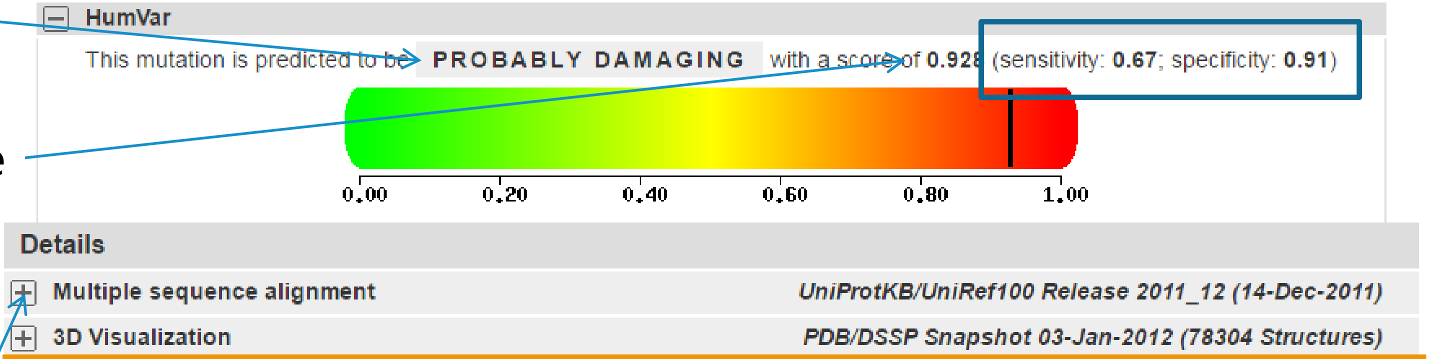
Output from Provean

PolyPhen2

- Assesses conservation and impact on the function of the region
- HumVar is trained on the difference between disease causing and neutral variants in humans

Prediction

Score



HumVar

This mutation is predicted to be **PROBABLY DAMAGING** with a score of **0.928** (sensitivity: **0.67**; specificity: **0.91**)

Details

Multiple sequence alignment	***UniProtKB/UniRef100 Release 2011_12 (14-Dec-2011)***
3D Visualization	***PDB/DSSP Snapshot 03-Jan-2012 (78304 Structures)***

Output from PolyPhen2

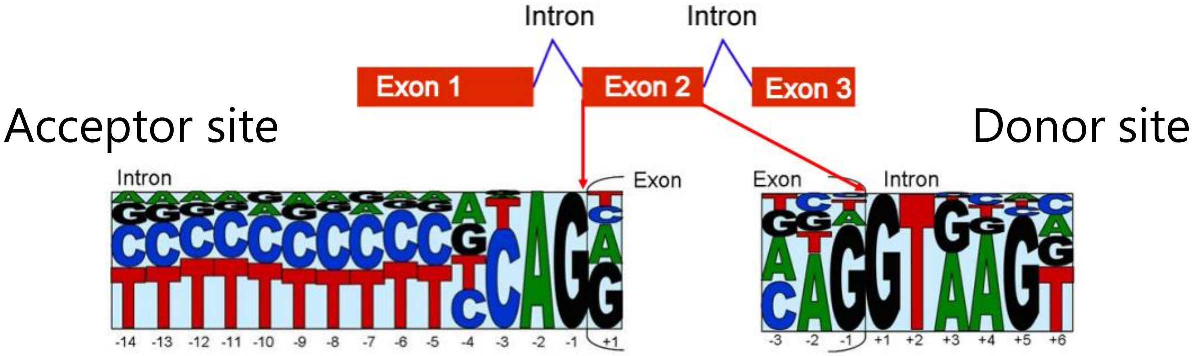
Alignment must be checked for validity

Your turn…

- Analyse the results from the in silico tools
  - What effect do they predict?
  - Are the results reliable?

Can we predict the effect?

- Splice predictors look for motif sequences and score them according to their “strength”
- If a signal at a natural site is reduced it could be skipped
- If a signal at a cryptic site is created/increased it could out compete the natural site and change the protein sequence



Canonical/consensus sequence for splice sites

Has anyone else studied it?

- Literature is a major source of information
- Functional/RNA studies are the holy grail of variant interpretation
- Disease specific studies can be useful but the information they contain is not always detailed
- Reliability is VERY important

Your turn...

- Use the search string provided in the form to perform a literature search
- You should use Google, Google Scholar and PubMed as a minimum
- Abstract information collection is acceptable but try to gage the reliability of any information you find

What does it all mean?

Sheffield Children's NHS
NHS Foundation Trust

Benign

- High population frequency
- Functional study
- Not linked to patient phenotype

Likely Benign

- Some frequency data
- In silico predicts neutral
- Not in a functional domain/likely to impact structure
- Few/no pathogenic mutations in the region

Unknown Significance

- Conflicting predictions from in silico tools
- No frequency data or literature

Likely Pathogenic

- No/low population frequency
- Suggestion that mutations in region have been identified before
- Pathogenic in silico predictions
- Disease specific literature

Pathogenic

- Segregates with the disease
- Functional study
- All evidence in agreement

Your turn...

- Combine all the lines of investigation and reach a consensus
- Use the ACMG classifier to see if there is agreement with your prediction
- Discuss your findings as a group and reach a concensus to feed back to the group