VASA: ACMG Variant Scoring Assistant: For Sequence Variant Classification, Variant Review and Teaching

Philip Davidson[1,2], Frances Smith[3], David Brawand[3] & Barnaby Clark[3]

Introduction

UK adoption of the American College of Medical Genetics and Genomics (ACMG) guidelines for genetic variant pathogenicity scoring initiated us to develop a tool to assist clinical scientists during this routine process.

VASA is a web-based tool that guides a user through the process of classifying variants against the ACMG categories; recording evidence that supports decisions and calculating the pathogenicity score. The guidelines can be overridden, allowing scientists to exercise judgment where necessary.

VASA retains the calculations, variant annotations and associated evidence in a repository for reference in future submissions. It does not however, store patient specific data, allowing sharing of evidence between users.

Variants may be scored multiple times and all scores and associated evidence are available to all users. Training VCF datasets can be uploaded allowing trainers to review how trainees scored variants, reviewing the evidence recorded and which ACMG categories caused any variation.

The Workflow

It is designed to be integrated into existing workflows. There are three stages to the process; importing variants, scoring, and interrogating the data:

Importing Variants

Variants can be imported into the system in several ways to create a ‘worklist’ for scoring:

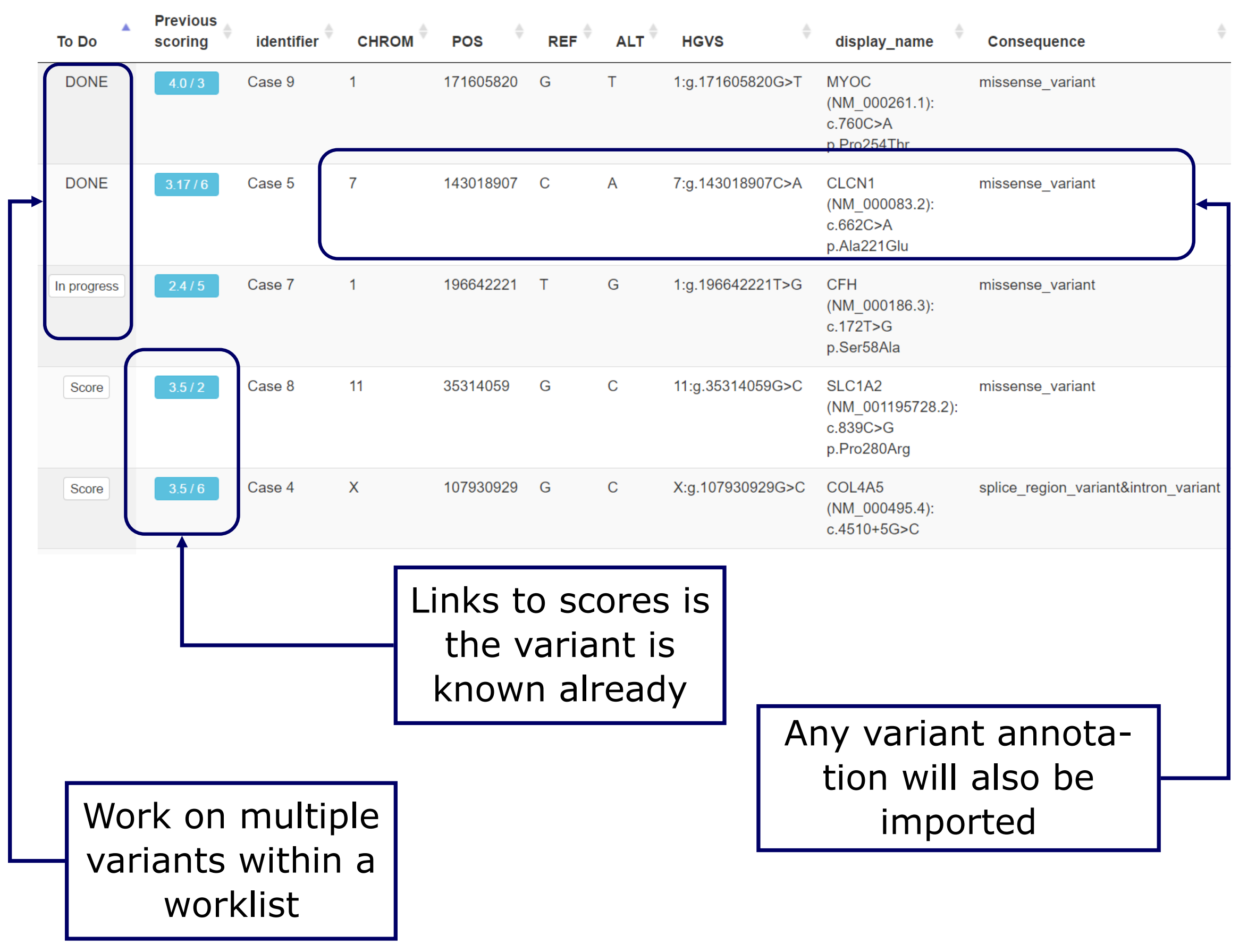
- Upload a VCF, with the option to use an annotated VCF file
- The tool’s API accepts the posting of variant lists, enabling a simple integration into existing workflows
- The system current accepts CSV files created by Sophia Genetics DDM®, but could be easily be modified to accept other bespoke data formats
- Individual variants can be manually entered for *ad hoc* scoring

Scoring

- The user can decide which variants they wish to score and the 28 categories are presented in a dashboard for the user to complete
- The final pathogenicity score is calculated as the scorer progresses
- The final score, as well as each weighting for each category, can be adjusted at the discretion of the scorer
- Pertinent annotation data will be displayed next to each category question
- The user will be prompted to include HPO terms with each score

Export/Query the Data

- Multiple scores can be stored against each variant
- All variants and scores can be interrogated and exported by API or by viewing in the web browser
- Scored variants can be exported to CSV
- All annotation and user submitted data is stored with the scores
- Date & time, user and organisation data is stored for easy auditing of variant scoring



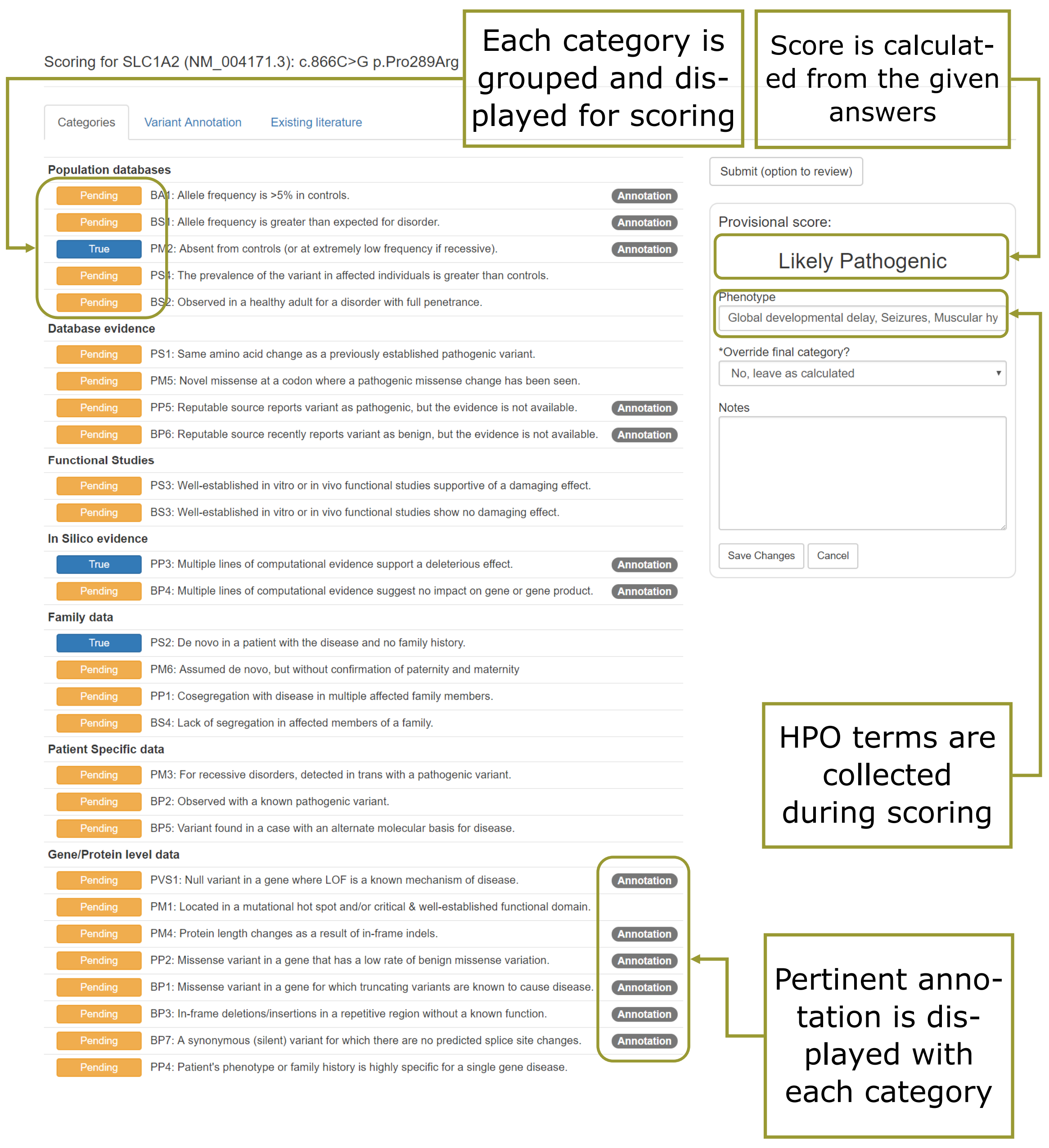
To Do	Previous scoring	identifier	CHROM	POS	REF	ALT	HGVS	display_name	Consequence
DONE	4.0 / 3	Case 9	1	171605820	G	T	1:g.171605820G>T	MYOC (NM_000261.1): c.760C>A p.Pro254Thr	missense_variant
DONE	3.17 / 6	Case 5	7	143018907	C	A	7:g.143018907C>A	CLCN1 (NM_000083.2): c.662C>A p.Ala221Glu	missense_variant
In progress	2.4 / 5	Case 7	1	196642221	T	G	1:g.196642221T>G	CFH (NM_000186.3): c.172T>G p.Ser58Ala	missense_variant
Score	3.5 / 2	Case 8	11	35314059	G	C	11:g.35314059G>C	SLC1A2 (NM_001195728.2): c.839C>G p.Pro280Arg	missense_variant
Score	3.5 / 6	Case 4	X	107930929	G	C	X:g.107930929G>C	COL4A5 (NM_000495.4): c.4510+5G>C	splice_region_variant&intron_variant

Links to scores is the variant is known already

Any variant annotation will also be imported

Work on multiple variants within a worklist

An example worklist



Scoring for SLC1A2 (NM_004171.3): c.866C>G p.Pro289Arg

Population databases

- Pending BA1: Allele frequency is >5% in controls. Annotation
- Pending BS1: Allele frequency is greater than expected for disorder. Annotation
- True PM2: Absent from controls (or at extremely low frequency if recessive). Annotation
- Pending PS4: The prevalence of the variant in affected individuals is greater than controls.
- Pending BS2: Observed in a healthy adult for a disorder with full penetrance.

Database evidence

- Pending PS1: Same amino acid change as a previously established pathogenic variant.
- Pending PM5: Novel missense at a codon where a pathogenic missense change has been seen.
- Pending PP5: Reputable source reports variant as pathogenic, but the evidence is not available. Annotation
- Pending BP6: Reputable source recently reports variant as benign, but the evidence is not available. Annotation

Functional Studies

- Pending PS3: Well-established in vitro or in vivo functional studies supportive of a damaging effect.
- Pending BS3: Well-established in vitro or in vivo functional studies show no damaging effect.

In Silico evidence

- True PP3: Multiple lines of computational evidence support a deleterious effect. Annotation
- Pending BP4: Multiple lines of computational evidence suggest no impact on gene or gene product. Annotation

Family data

- True PS2: De novo in a patient with the disease and no family history.
- Pending PM6: Assumed de novo, but without confirmation of paternity and maternity
- Pending PP1: Cosegregation with disease in multiple affected family members.
- Pending BS4: Lack of segregation in affected members of a family.

Patient Specific data

- Pending PM3: For recessive disorders, detected in trans with a pathogenic variant.
- Pending BP2: Observed with a known pathogenic variant.
- Pending BP5: Variant found in a case with an alternate molecular basis for disease.

Gene/Protein level data

- Pending PVS1: Null variant in a gene where LOF is a known mechanism of disease. Annotation
- Pending PM1: Located in a mutational hot spot and/or critical & well-established functional domain.
- Pending PM4: Protein length changes as a result of in-frame indels. Annotation
- Pending PP2: Missense variant in a gene that has a low rate of benign missense variation. Annotation
- Pending BP1: Missense variant in a gene for which truncating variants are known to cause disease. Annotation
- Pending BP3: In-frame deletions/insertions in a repetitive region without a known function. Annotation
- Pending BP7: A synonymous (silent) variant for which there are no predicted splice site changes. Annotation
- Pending PP4: Patient's phenotype or family history is highly specific for a single gene disease.

Submit (option to review)

Provisional score: Likely Pathogenic

Phenotype: Global developmental delay, Seizures, Muscular hy

*Override final category?: No, leave as calculated

Notes

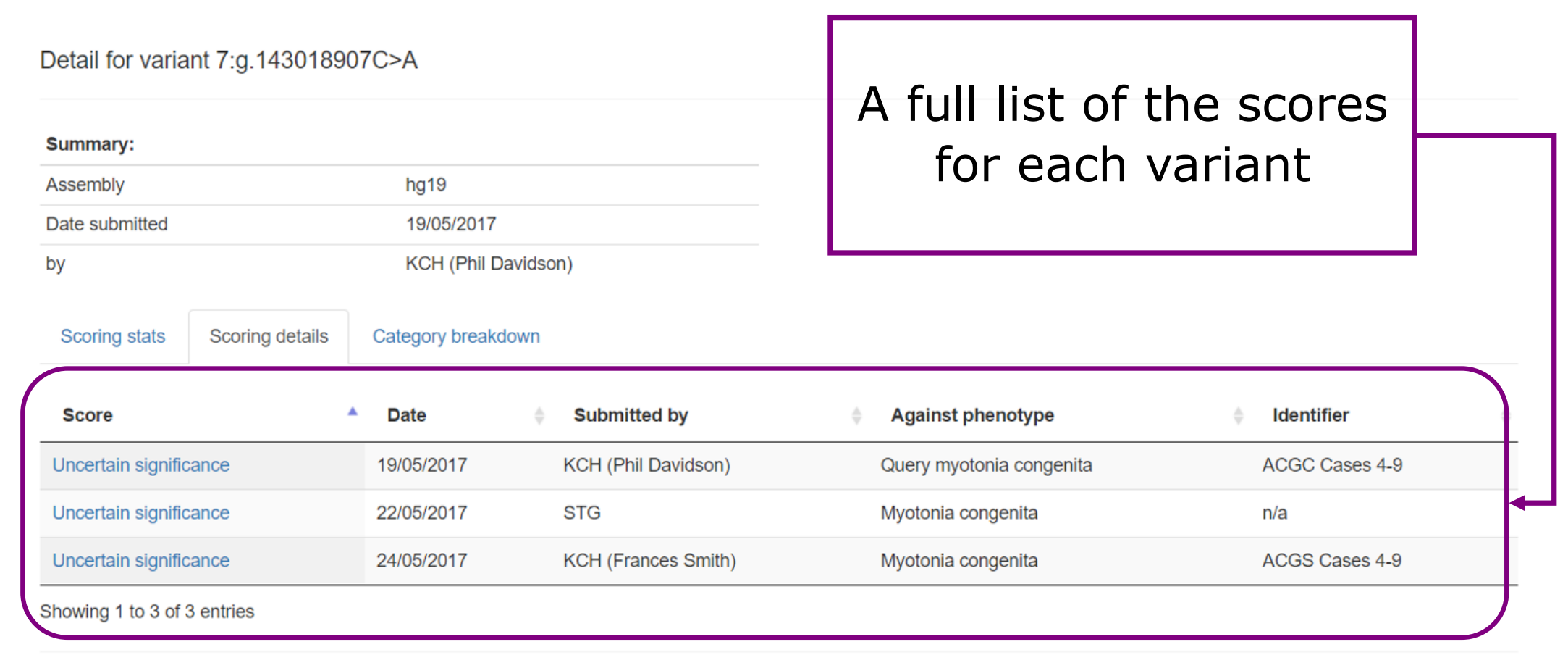
Each category is grouped and displayed for scoring

Score is calculated from the given answers

HPO terms are collected during scoring

Pertinent annotation is displayed with each category

The scoring dashboard



Detail for variant 7:g.143018907C>A

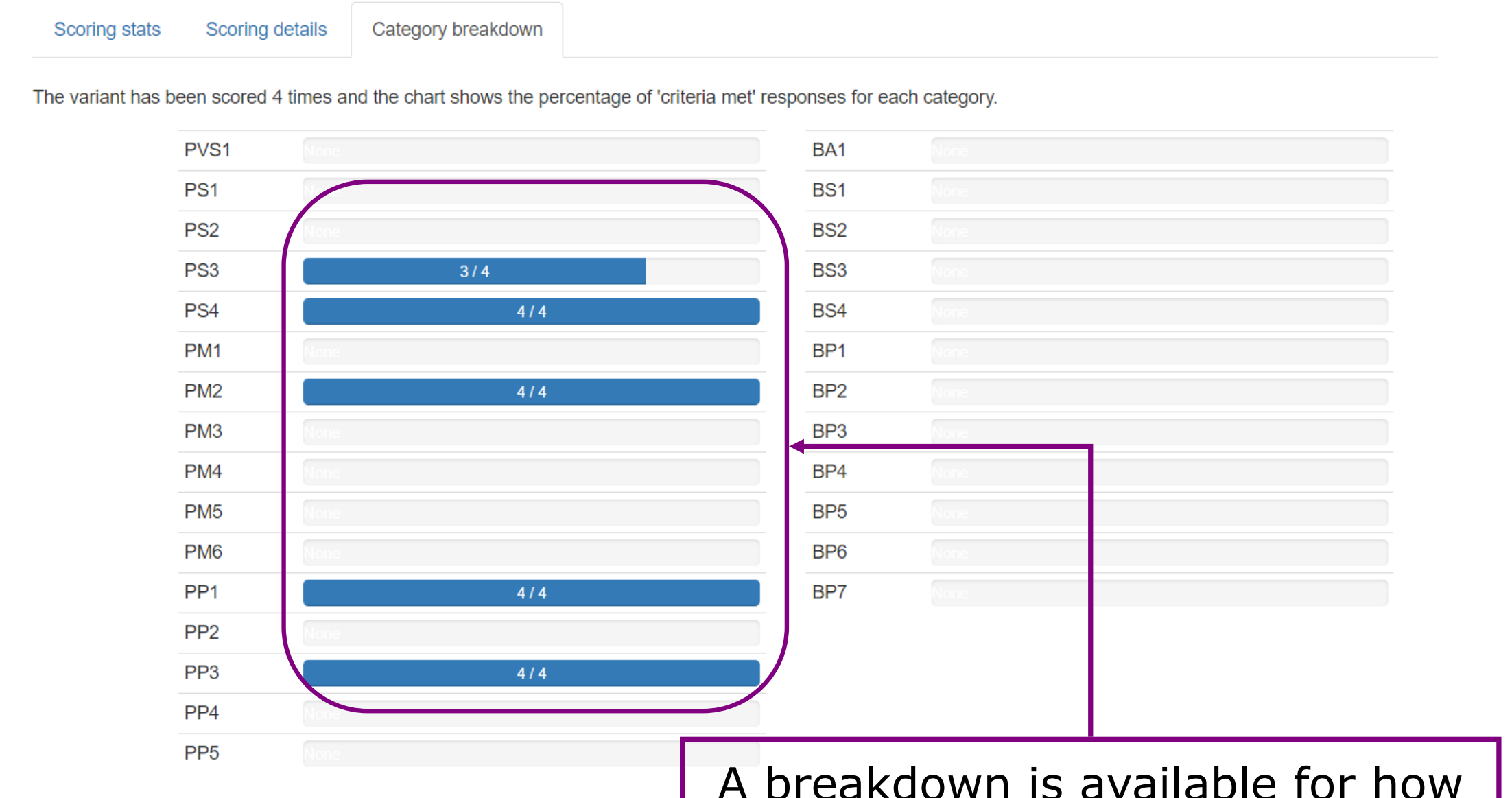
Summary:	
Assembly	hg19
Date submitted	19/05/2017
by	KCH (Phil Davidson)

Score	Date	Submitted by	Against phenotype	Identifier
Uncertain significance	19/05/2017	KCH (Phil Davidson)	Query myotonia congenita	ACGC Cases 4-9
Uncertain significance	22/05/2017	STG	Myotonia congenita	n/a
Uncertain significance	24/05/2017	KCH (Frances Smith)	Myotonia congenita	ACGS Cases 4-9

Showing 1 to 3 of 3 entries

A full list of the scores for each variant

Variant data



The variant has been scored 4 times and the chart shows the percentage of 'criteria met' responses for each category.

A breakdown is available for how each category has been marked across all scores

Category breakdowns

Technical Specifications:

- Built using Django 1.11 & Python 3.5.2
- The app is database, web server & OS independent
- Currently tested on Windows with IIS and MySQL & MS SQL Server, and CentOS, Apache with MySQL

Try me:

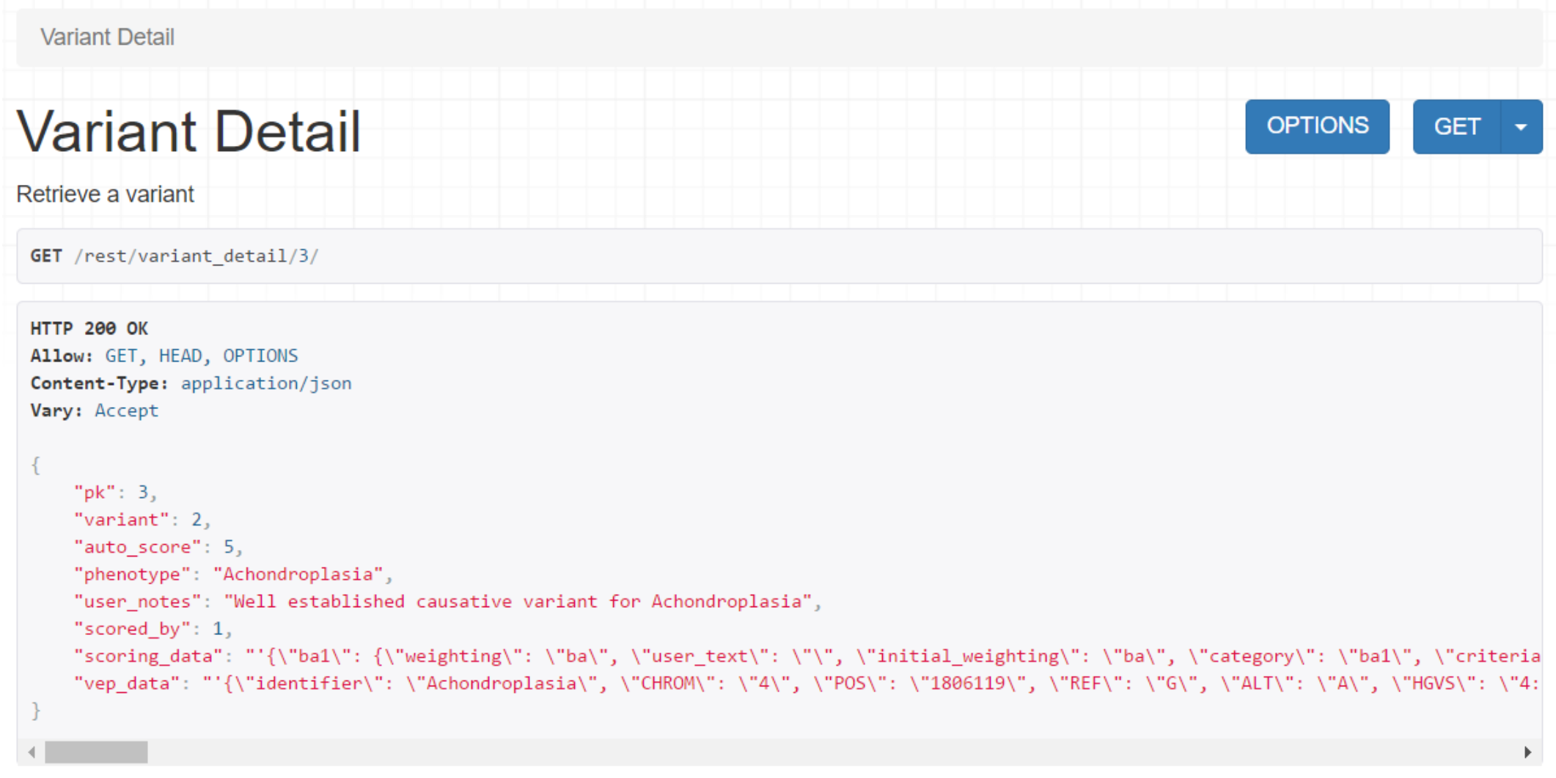
http://li446-39.members.linode.com

A Training Tool

VASA is an ideal tool for training as well as integration in day-to-day workflows.

It can hold training sets of variants that an be scored by multiple users. This will allow the auditing of how the ACMG guidelines are being applied across organisations, and see which areas are discordant.

The NEQAS variant interpretation scheme may benefit from a tool like this as it allows instant review of how variants have been scored and interpreted. If adopted more widely it would also identify variants that could be used in the scheme.



Variant Detail

Variant Detail

Retrieve a variant

```
GET /rest/variant_detail/3/
```

```
HTTP 200 OK
Allow: GET, HEAD, OPTIONS
Content-Type: application/json
Vary: Accept

{
    "pk": 3,
    "variant": 2,
    "auto_score": 5,
    "phenotype": "Achondroplasia",
    "user_notes": "Well established causative variant for Achondroplasia",
    "scored_by": 1,
    "scoring_data": "'{\"ba1\": {\"weighting\": \"ba\", \"user_text\": \"\", \"initial_weighting\": \"ba\", \"category\": \"ba1\", \"criteria
    "vep_data": "'{\"identifier\": \"Achondroplasia\", \"CHROM\": \"4\", \"POS\": \"1806119\", \"REF\": \"G\", \"ALT\": \"A\", \"HGVS\": \"4:
}
```

The API has been build with the Django REST Framework

[1] Faculty of Medical and Human Sciences, University of Manchester, Oxford Road, Manchester, UK
[2] South West Thames Regional Genetics Service, St George's University Hospitals NHS Foundation Trust, London
[3] Department of Molecular Pathology, Viapath at King's College Hospital NHS Foundation Trust, London