CardioClassifier – demonstrating the power of disease- and gene-specific computational decision support for clinical genome interpretation

Nicola Whiffin, PhD1,2,3,†,Roddy Walsh, MSc1,2,Risha Govind, MSc1,2,Matthew Edwards, MSc4,Mian Ahmad, PhD1,2,Xiaolei Zhang, MSc1,2,Upasana Tayal, BMBCh, MRCP1,2,Rachel Buchan, MSc1,2,William Midwinter, BSc1,2,Alicja E Wilk, BSc1,2,Hanna Najgebauer, PhD1,2,Catherine Francis, MA, MRCP1,2,Sam Wilkinson, BSc4,Thomas Monk, MSc4,Laura Brett, MPhil4,Declan P O'Regan, PhD, FRCR1,2,3,†,Sanjay K Prasad, MD, FRCP1,2,Deborah J Morris-Rosendahl, PhD1,4,Paul JR Barton, PhD1,2,Elizabeth Edwards, PhD1,2,James S Ware, PhD, MRCP1,2,3,\*,Stuart A Cook, PhD, MRCP1,2,5,6,\*

1National Heart & Lung Institute, Imperial College London  
2Cardiovascular Research Centre at Royal Brompton and Harefield NHS Foundation Trust, London, UK  
3MRC London Institute of Medical Sciences, Imperial College London  
4Clinical Genetics and Genomics Laboratory, Royal Brompton and Harefield NHS Foundation Trust, London, UK  
5National Heart Centre Singapore, Singapore  
6Duke-National University of Singapore, Singapore

\*These authors jointly supervised this work  
†Corresponding author

**SUPPLEMENTARY INFORMATION**

## SUPPLEMENTARY NOTE 1 - Parameterisation of ACMG/AMP rules

CardioClassifier currently includes 11 ICCs for which we have solid estimates of both prevalence and heterogeneity in order to generate accurate maximum cut-offs for allele frequency filtering (activation of BS1/PM2) (<http://www.nature.com/gim/journal/vaop/ncurrent/full/gim201726a.html>). Genes selected for inclusion have definitive links to these ICCs, defined as one or more of the following:

* Truncating variants in this gene activate PVS1 (method for determining this is outlined below)
* The gene shows a significant excess of either truncating or non-truncating variants in disease cohorts over controls (<http://www.nature.com/gim/journal/vaop/ncurrent/full/gim201690a.html>)
* Individual variants in this gene are significantly enriched in disease cohorts over controls (i.e. at least one variant activates PS4)
* The gene is common included on gene testing panels found in UKGTN and there are multiple pathogenic assertions in ClinVar for variants in this gene

Moving forwards, genes identified as strongly/definitively associated with an ICC by the ClinGen Cardiovascular Domain Working Group will be added as they can be parameterized.

### Rules activated consistently across all genes/diseases

The following rules are activated consistently, irrespective of the disease and gene analysed:

*PS1 – same amino acid change as previously established pathogenic variant*

Activated when a variant causing the same amino acid change is categorised as ‘Pathogenic’ for the phenotype of interest in ClinVar by multiple submitters with no conflicting evidence. If data is from only a single submitter, it is displayed on the report but is not used to activate PS1.

*PM4 – protein length changes as a result of in-frame deletions/insertions in a non-repeat region or stop-lost variants*

Activated if the variant is annotated as ‘stop-lost’ or ‘in-frame indel’ by the Ensembl Variant Effect Predictor (VEP). For in-frame indels the rule is only activated if the variant is not within a region annotated by repeat masker.

*PM5 – novel missense change at an amino acid residue where a different missense change has previously been determined to be pathogenic*

Activated when a different missense variant at the same residue is categorised as ‘Pathogenic’ for the phenotype of interest in ClinVar by multiple submitters with no conflicting evidence. If data is from only a single submitter, it is displayed on the report but is not used to activate PM5.

*PP3 – multiple lines of computation evidence support a deleterious effect on the gene or gene product*

The following eight prediction algorithms are used: SIFT, PolyPhen2 var, LRT, Mutation Taster, Mutation Assessor, FATHMM, CADD and Grantham scores. The rule is activated when the following conditions are met:

1. At least five tools predict deleterious, with at most 1 tool predicting benign, and <3 with “unknown” classifications

*or*

1. If >=3 tools have unknown outcomes, all other tools predict damaging

*BA1 – allele frequency is >5%*

Due to the rarity of our disorders this threshold was reduced and this rule is activated for all variants with a filtering allele frequency (<http://www.nature.com/gim/journal/vaop/ncurrent/full/gim201726a.html>) >0.1% for heterozygotes or >3.16% (sqrt(0.001)) for homozygotes in the exome aggregation consortium (ExAC) database. We also use a set of internal healthy volunteers from the UK to detect Illumina platform specific errors. As this is a small dataset we use a threshold of 5% and activate BA1 if a variant is rare/absent from ExAC but present in >5% of healthy volunteers.

*BP3 - in-frame deletions/insertions in a repetitive region without known function*

Activated for any in-frame insertion or deletion that falls within a region annotated by repeat masker.

*BP4 - multiple lines of computational evidence suggest no impact on gene or gene product*

The following eight prediction algorithms are used: SIFT, PolyPhen2 var, LRT, Mutation Taster, Mutation Assessor, FATHMM, CADD and Grantham scores. The rule is activated when the following conditions are met:

1. At least five tools predict benign, with at most 1 tool predicting deleterious, and <3 with “unknown” classifications

*or*

1. If >=3 tools have unknown outcomes, all other tools predict benign

### Disease/gene specific rules

The following rules are activated for certain genes or diseases where data was available for analysis:

*PVS1 - null variant in a gene where Loss of Function (LoF) is a known mechanism of disease*

Activated for gene/disease pairs that meet one or more of the following conditions:

1. Genes with a significant burden of truncating variants in cases against controls from analysis of 7,855 cardiomyopathy cases and 60,706 controls (<http://biorxiv.org/content/early/2016/02/24/041111>)
2. Genes with an excess of truncating variants in HGMD over ExAC
3. Phenocopy genes where 2. is true

PVS1 is therefore activated for the following gene-disease pairs:

|  |  |  |
| --- | --- | --- |
| Disease | Gene | Condition met |
| DCM | *LMNA* | 1 |
| DCM | *DSP* | 1 |
| DCM | *VCL* | 1 |
| DCM | *BAG3* | 2 |
| HCM | *MYBPC3* | 1 |
| HCM | *TNNT2* | 1 |
| HCM | *PLN* | 1 |
| HCM | *FHL1* | 2 |
| HCM | *GLA* | 3 |
| HCM | *LAMP2* | 3 |
| ARVC/D | *DSP* | 1 |
| ARVC/D | *DSG2* | 1 |
| ARVC/D | *DSC2* | 1 |
| ARVC/D | *PKP2* | 1 |
| ncCM | *TAZ* | 2 |
| LQTS | *KCNQ1* | 2 |
| LQTS | *KCNH2* | 2 |
| Brugada syndrome | *SCN5A* | 2 |
| Marfan syndrome | *FBN1* | 2 |
| FH | *LDLR* | 2 |

NB: For this rule, truncating variants were taken as a proxy for variants that cause LoF. Care must be taken to ensure confidence that the given variant does indeed cause LoF.

*PVS1\_strong - TTN truncating variant in a constitutively expressed exon*

This rule was created specifically for analysing variants in the TTN gene where truncating variants are only known to cause disease if located in exons that are constitutively expressed in the heart (i.e. PVS1 doesn’t always apply). It is activated within a DCM test for nonsense, frameshift and essential splice site variants in TTN within exons with proportion spliced in (PSI) >0.9. Details of PSI calculation can be found in: <http://www.ncbi.nlm.nih.gov/pubmed/25589632>.

*PS4 - the prevalence of the variant in affected individuals is significantly increased compared with controls*

This rule has been assessed for DCM, HCM, ARVD/C, Brugada syndrome and LQTS using variants from the following publications:

|  |  |
| --- | --- |
| Disease | Publication |
| DCM, HCM, ARVD/C | <http://www.nature.com/gim/journal/vaop/ncurrent/full/gim201690a.html> |
| LQTS | <https://www.ncbi.nlm.nih.gov/pubmed/19716085> |
| Brugada syndrome | <http://www.ncbi.nlm.nih.gov/pubmed/20129283> |

A Fisher’s exact test was used to test for association against the ExAC dataset. If a variant does not exist in ExAC, the total count of individuals is taken as the proportion covered at >=15x at the variant position. A Bonferroni correction based on the number of rare variants (freq <0.0001) across all tested genes in either ExAC or the case cohort, was used to assess statistical significance. This rule is activated if the count in the cases was >2 and the Fisher’s *P*<1.79x10-6.

*PM1 - located in a mutational hot-spot and/or critical and well-established functional domain*

The following gene-disease pairs have been analysed for this rule with data obtained from the reference in the table:

|  |  |  |
| --- | --- | --- |
| Disease | Gene(s) | Publication |
| LQTS | *KCNQ1*, *KCNH2*, *SCN5A* | <https://www.ncbi.nlm.nih.gov/pubmed/19716085> |
| Brugada syndrome | *SCN5A* | <http://www.ncbi.nlm.nih.gov/pubmed/20129283> |
| HCM | *MYH7* | <http://www.nature.com/gim/journal/vaop/ncurrent/full/gim201690a.html> |
| CPVT | *RYR2* | <http://www.ncbi.nlm.nih.gov/pubmed/199260154> |
| DCM | *RBM20* | <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2782634/> |
| Marfan syndrome | *FBN1* | <http://www.cell.com/ajhg/fulltext/S0002-9297(07)62603-2> |

For RYR2 the regions used are directly taken from the reference, these are the ‘N terminal hot-spot’, the ‘Central hot-spot’ and the ‘Channel region hot-spot’. For *RBM20*, PM1 is activated for missense mutations in the 9 amino acid hotspot identified by Brauch et al. For *FBN1*, EF-like domains were identified using <http://www.umd.be/FBN1/> and PM1 is activated for missense mutations affecting a cysteine residue.

For the other gene-disease pairs, variant frequencies from the references were compared with frequencies in the ExAC dataset over a set of defined regions. For the arrhythmia genes, regions were previously defined protein domains and the highly conserved regions of the *KCNQ1* C-terminus <http://www.ncbi.nlm.nih.gov/pubmed/25854863>. For MYH7, clusters were calculated from the data itself.

This rule was activated if the variant was within a region with etiological fraction (i.e. the estimated proportion of cases with a variant where the variant is causative, essentially equivalent to a prior probability of pathogenicity) ≥0.95 with Fisher’s exact P<0.05. Variants in regions with 0.75<EF<0.5 activate PP2 as described below.

*PM2 - absent/low frequency in controls*

Activated if the ExAC filtering allele frequency is less than the calculated maximum tolerated allele frequency for the disease. Details on how the maximum tolerated AF and filtering AF are determined can be found here: <http://www.nature.com/gim/journal/vaop/ncurrent/full/gim201726a.html>

NB: be aware that population data for insertions/deletions may be poorly covered. We do not currently check there is coverage at positions where no variant is found.

*BS1 - allele frequency is greater than expected for disorder*

Activated if the ExAC filtering allele frequency is greater the calculated maximum tolerated allele frequency for the disease. Details on how the maximum tolerated AF and filtering AF are determined can be found here: <http://www.nature.com/gim/journal/vaop/ncurrent/full/gim201726a.html>

*PP2 – missense variant in a gene with a low rate of benign missense variation and in which missense variants are a common mechanism of disease*

We use PP2 to activate both gene regions with lower confidence than PM1 (i.e. EF ≥0.75) or genes that are both constrained and where missense variants are commonly disease causing.

For the LQTS, Brugada syndrome and HCM genes in PM1, PP2 is activated where EF ≥0.75 and Fisher’s exact P<0.05.

For all other genes involved in DCM, HCM and ARVD/C, PP2 activation is based on data from <http://www.nature.com/gim/journal/vaop/ncurrent/full/gim201690a.html> and the ExAC database for genes with missense etiological fraction (i.e. the estimated proportion of cases with a variant where the variant is causative, essentially equivalent to a prior probability of pathogenicity) ≥0.75 with Fisher’s exact P<0.05.

Finally, we activate PP2 for two additional genes that are constrained in ExAC (missense Z score > 3.09) and where missense mutations are a common disease mechanism. These are FBN1 in Marfan syndrome and PTPN11 for Noonan syndrome.

As we see PP2 as a lower confidence version of PM1, this rule is never activated if PM1 is already used.

**Paralogue annotation**

For genes with paralogues that are known to be involved in disease, variants at equivalent residues in these paralogues can identify residues that are likely intolerant to variation and hence may be more likely to harbour disease causing variants. We use previously published data (<http://jmg.bmj.com/content/early/2013/10/17/jmedgenet-2013-101917.full>) to activate two rules relating to variants at equivalent residues and the same amino acid change at these residues. Only variants with high-confidence mapping to a paralogous gene are used (M-coffee mapping score >3 and the same reference amino acid). These rules are relevant to all genes currently analysed for LQTS, Brugada syndrome and CPVT. Further information is available at <http://cardiodb.org/Paralogue_Annotation/>.

*PS1\_moderate - Equivalent amino acid change as an established pathogenic variant in a paralogous gene*

*PM5\_supporting - Missense change at an amino acid residue where a pathogenic missense change has been seen in the equivalent residue of a paralogous gene*

As we use these rules in the same way as PS1 and PM5, but treating evidence in paralogues as weaker than evidence in the same gene, these rules are never activated if either PS1 or PM5 are already active.

NB: As this data currently uses variants ‘reported’ in HGMD as pathogenic that have not been manually curated, it is important to verify whether the report of pathogenicity for the observation in the paralogue is robust.

### User only rules

The following rules represent clinical data or data that requires manual curation, and hence are not currently computationally predicted/activated for any genes or diseases. These rules will be activated if data has been manually entered into the CardioClassifier knowledge base, or can be activated interactively at the user's discretion.

|  |  |
| --- | --- |
| Rule | Description |
| PS2 | *de novo* (both maternity and paternity confirmed) in a patient with disease and no family history |
| PS3 | well-established *in vitro* or *in vivo* functional studies supportive of a damaging effect |
| PM3 | for recessive disorders, detected in trans with a pathogenic variant |
| PM6 | assumed *de novo*, but without confirmation of paternity and maternity |
| PP1 | co-segregation with disease in multiple affected family members in a gene definitively known to cause disease |
| PP4 | patient’s phenotype or family history is highly specific for a disease with a single genetic aetiology |
| PP5 | reputable source recently reports a variant as pathogenic, but the evidence is not available to perform an independent evaluation |
| BS2 | observed in a healthy individual 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 |
| 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 penetrance dominant gene/disorder or observed in cis with a pathogenic variant in any inheritance pattern |
| BP5 | variant found in a case with an alternative molecular basis for disease |
| BP6 | reputable source recently reports variant as benign, but the evidence is not available to perform an independent evaluation |

In addition, the tool does not currently annotate synonymous variants so rule BP7 is not utilised.

## SUPPLEMENTARY NOTE 2

### Benchmarking against manually curated variants

As part of the developmental process, we compared CardioClassifier to an orthogonally curated dataset of 44 protein-altering variants from patients referred to the Clinical Genetics and Genomics Laboratory at the Royal Brompton and Harefield NHS Trust, that had been manually curated according to the ACMG/AMP guidelines as part of routine clinical service.

Of 153 data points activated by the clinical lab, 19 represented case-level/functional data which we would not expect CardioClassifier to retrieve. Of the remaining 134, 79 (59.0%) were consistently activated. On review of the 55 discrepancies, in 12 cases the clinical lab revised the evidence (due to inconsistencies or inappropriate combinations of rules), in 3 cases, the rule was updated within CardioClassifier, and in 40 cases we agreed to differ. For these rules, CardioClassifier is set with a more conservative threshold, leaving a user to activate it if they feel it is appropriate.

## SUPPLEMENTARY TABLES

|  |  |  |  |
| --- | --- | --- | --- |
| Disease | Prevalence | Max Allelic Contribution | Maximum population AF |
| DCM | 1/250 | 0.021 | 8.4e-05 |
| HCM | 1/500 | 0.02 | 4e-05 |
| ARVD/C | 1/1,000 | 0.092 | 9.2e-05 |
| RCM | 1/5,000 | 0.1 | 2e-05 |
| ncCM | 1/100 | 0.1 | 0.001 |
| Noonan syndrome | 1/1,000 | 0.1 | 1e-04 |
| Long QT syndrome | 1/2,000 | 0.016 | 8e-06 |
| Brugada syndrome | 1/1,000 | 0.01 | 1e-05 |
| CPVT | 1/10,000 | 0.1 | 5e-05 |
| Marfan syndrome | 1/3,000 | 0.015 | 5e-06 |
| FH | 1/250 | 0.1 | 4e-04 |

**Supplementary Table S1: Details of allele frequency cutoffs for each disease parameterised in CardioClassifier.** Shown is the reported prevalence of each condition, the maximum proportion of disease attributable to a single variant and the resulting maximum population frequency. Full details of frequency calculations can be found in Whiffin et al. GiM 2017.

Table continues below

|  |  |  |
| --- | --- | --- |
|  | ClinGen activated rules |  |
| Variant | Computational | Case-level |
| c.428G>A | PS4, PM2 | PP1 |
| c.732+1G>A | PM2, PVS1\_mod\*, PP3 | - |
| c.788T>C | PS4, PM1, PM2, PP3 | PP1\_strong |
| c.1106G>A | PM1, PM2 | PM6, PS4\_moderate |
| c.1157A>G | PM1, PM2, PP3 | PM6 |
| c.1207C>T | PS4, PM1, PM2, PM5, PP3 | PP1\_strong |
| c.1208G>A | PS4, PM1, PM2, PM5, PP3 | PS3, PP1\_strong |
| c.1357C>T | PS4, PM1, PM2, PP3 | PS2, PS3, PP1\_strong |
| c.1358G>A | PS4, PM1, PM2, PM5, PP3 | PM6 |
| c.1477\_1478delAT | PM2, PVS1\_mod\*, PP3 | - |
| c.1594T>C | PM1, PM2, PP3 | PS3, PP1\_strong, PS4\_supporting |
| c.1750G>C | PS4, PM1, PM2, PP3 | PP1\_moderate |
| c.1988G>A | PS4, PM1, PM2 | PP1\_strong |
| c.2146G>A | PS4, PM1, PM2, PP3 | PP1\_strong, PM6 |
| c.2155C>T | PS4, PM1, PM2, PM5, PP3 | PS2, PS3, PP1\_strong |
| c.2156G>A | PS4, PM1, PM2, PM5, PP3 | PP1\_strong |
| c.2162+4G>A | BA1, BP4\* | - |
| c.2167C>G | PS4, PM1, PM2, PM5, PP3 | PP1\_strong |
| c.2167C>T | PS4, PM1, PM2, PM5, PP3 | PP1\_strong, PM6, BP2 |
| c.2207T>C | PS4, PM1, PM2 | PP1\_moderate |
| c.2221G>C | PS4, PM1, PM2, PM5, PP3 | PM6, PP1 |
| c.2221G>T | PS4, PM1, PM2, PM5, PP3 | PP1\_moderate |
| c.2360G>A | PM1 | BP2 |
| c.2513C>T | PM1, PM2, PP3 | PS2, PS4\_supporting |
| c.2539\_2541delAAG | PS4, PM2, PM4 | PP1 |
| c.2539A>G | PS4, PM1, PM2, PP3 | - |
| c.2608C>T | PM1, PM2, PM5, PP3 | PS4\_supporting |
| c.2609G>A | PS4, PM1, PM2 | PP1\_strong |
| c.2678C>T | PM1, PM2 | PP1\_moderate, PS4\_supporting, BS4 |
| c.2681A>G | PS4, PM1, PM2, PP3 | PP1\_moderate |
| c.2717A>G | PS4, PM1, PM2 | PP1\_strong |
| c.2722C>G | PS4, PM1, PM2, PP3 | PP1\_strong |
| c.2791\_2793delGAG | PM2, PM4 | PP1\_moderate, PS4\_supporting |
| c.3133C>T | PS4, PM2, PP3 | - |
| c.3158G>A | PS4, PM2, PP3 | PP1\_moderate |
| c.3169G>A | PM2, PP3 | PS4\_moderate |
| c.3286G>T | PP3 | - |
| c.3382G>A | - | - |
| c.3578G>A | PP3 | - |
| c.3658\_3660delGAG | PM2, PM4 | PP1, PS4\_supporting |
| c.3981C>A | PP3 | BS4 |
| c.4066G>A | PS4, PM2, PP3 | PP1\_moderate |
| c.4130C>T | PS4, PM2, PP3 | - |
| c.4258C>T | PM2, PP3 | PS4\_moderate |
| c.4276G>A | PM2, PP3 | PS4\_supporting |
| c.4377G>T | PP3, BS1 | - |
| c.4588C>T | PM2, PVS1\_mod\* | - |
| c.4909G>A | - | - |
| c.5135G>A | PS4, PM2, PP3 | PP1 |
| c.5302G>A | PM2, PP3 | PS4\_supporting |
| c.5326A>G | PP3 | - |
| c.5329G>A | PM2, PP3 | - |
| c.5401G>A | PM2, PP3 | PM6, PS4\_supporting |
| c.5588G>A | PM2, PP3 | PS4\_supporting |
| c.5704G>C | - | - |
| c.5726G>C | PM2, PP3 | PM6 |
| c.5740G>A | PM2, PP3 | PM6, PS4\_supporting |

Table continues below

|  |  |
| --- | --- |
|  | CardioClassifier activated rules |
| Total (computational) | Automatic |
| 3 (2) | PS4, PM2, PP2 |
| 3 (3) | PM2 |
| 5 (4) | PM1, PM2, PP3 |
| 4 (2) | PM1, PM2 |
| 4 (3) | PM1, PM2, PP3 |
| 6 (5) | PS4, PM1, PM2, PM5, PP3 |
| 7 (5) | PS4, PM1, PM2, PM5 |
| 7 (4) | PS4, PM1, PM2, PP3 |
| 6 (5) | PM1, PM2, PM5, PP3 |
| 3 (3) | PM2 |
| 6 (3) | PM1, PM2, PP3 |
| 5 (4) | PS4, PM1, PM2, PP3 |
| 4 (3) | PS4, PM1, PM2 |
| 6 (4) | PS4, PM1, PM2, PP3 |
| 7 (5) | PS4, PM1, PM2, PM5, PP3 |
| 6 (5) | PS4, PM1, PM2, PM5 |
| 2 (2) | BA1 |
| 6 (5) | PM1, PM2, PM5, PP3 |
| 8 (5) | PS4, PM1, PM2, PM5, PP3 |
| 4 (3) | PS4, PM1, PM2 |
| 7 (5) | PM1, PM2, PM5, PP3 |
| 6 (5) | PS4, PM1, PM2, PP3 |
| 2 (1) | PM1 |
| 5 (3) | PM1, PM2, PP3 |
| 4 (3) | PM2, PM4 |
| 4 (4) | PS4, PM1, PM2 |
| 5 (4) | PM1, PM2, PM5, PP3 |
| 4 (3) | PS4, PM1, PM2, PP3 |
| 5 (2) | PM1, PM2 |
| 5 (4) | PS4, PM1, PM2, PP3 |
| 4 (3) | PS4, PM1, PM2 |
| 5 (4) | PS4, PM1, PM2, PP3 |
| 4 (2) | PM2, PM4 |
| 3 (3) | PM2, PP2, PP3 |
| 4 (3) | PM2, PP2, PP3 |
| 3 (2) | PM2, PP2, PP3 |
| 1 (1) | PP2, PP3 |
| 0 (0) | PP2 |
| 1 (1) | PM2, PP2, PP3 |
| 4 (2) | PM2, PM4 |
| 2 (1) | PP3 |
| 4 (3) | PS4, PM2 |
| 3 (3) | PS4, PM2 |
| 3 (2) | PM2, PP3 |
| 3 (2) | PM2, PP3 |
| 2 (2) | BS1 |
| 2 (2) | PM2 |
| 0 (0) | - |
| 4 (3) | PS4, PM2 |
| 3 (2) | PM2, PP3 |
| 1 (1) | - |
| 2 (2) | PM2 |
| 4 (2) | PM2, PP3 |
| 3 (2) | PM2 |
| 0 (0) | - |
| 3 (2) | PM2, PP2, PP3 |
| 4 (2) | PM2, PP2, PP3 |

Table continues below

|  |  |  |  |
| --- | --- | --- | --- |
|  |  | Retrieved rules |  |
| Through links | Link source | Computational | Case-level |
| PP1 | ClinVar | 2 | 1 |
| - | - | 1 | - |
| PP1\_strong | ClinVar, Google | 3 | 1 |
| PM6, PS4\_moderate | ClinVar | 2 | 2 |
| PM6 | ClinVar | 3 | 1 |
| PP1\_strong | ClinVar | 5 | 1 |
| PS3, PP1\_strong | ClinVar, Google | 4 | 2 |
| PS2, PS3, PP1\_strong | ClinVar, Google | 4 | 3 |
| PM6 | ClinVar | 4 | 1 |
| - | - | 1 | - |
| PS3, PP1\_strong | ClinVar, Google | 3 | 2 |
| PP1\_moderate | ClinVar | 4 | 1 |
| PP1\_strong | ClinVar | 3 | 1 |
| PP1\_strong, PM6 | ClinVar | 4 | 2 |
| PS2, PS3, PP1\_strong | ClinVar, Google | 5 | 3 |
| PP1\_strong | ClinVar | 4 | 1 |
| - | - | 1 | - |
| PP1\_strong | ClinVar | 4 | 1 |
| PP1\_strong, PM6, BP2 | ClinVar | 5 | 3 |
| PP1\_moderate | ClinVar | 3 | 1 |
| PM6, PP1 | ClinVar | 4 | 2 |
| PP1\_moderate | ClinVar | 4 | 1 |
| BP2 | ClinVar | 1 | 1 |
| PM6 | ClinVar | 3 | 0 |
| PP1 | ClinVar | 2 | 1 |
| - | - | 3 | - |
| PS4\_supporting | ClinVar | 4 | 1 |
| PP1\_strong | ClinVar | 3 | 1 |
| PS4\_supporting, PP1 supporting |  | 2 | 1 |
| PP1\_moderate | ClinVar | 4 | 1 |
| PP1\_strong | ClinVar | 3 | 1 |
| PP1\_strong | ClinVar | 4 | 1 |
| PP1\_moderate, PS4\_supporting | ClinVar, Google | 2 | 2 |
| - | - | 2 | - |
| PP1\_moderate | - | 2 | 1 |
| PS4\_supporting | ClinVar | 2 | 0 |
| - | - | 1 | 0 |
| - | - | - | - |
| - | - | 1 | - |
| PP1 | ClinVar | 2 | 1 |
| BS4 | ClinVar | 1 | 1 |
| - | - | 2 | 0 |
| - | - | 2 | - |
| PS4\_moderate | ClinVar | 2 | 1 |
| - | - | 2 | 0 |
| - | - | 1 | - |
| - | - | 1 | - |
| - | - | - | - |
| PP1 | ClinVar | 2 | 1 |
| PS4\_supporting | ClinVar | 2 | 1 |
| - | - | 0 | - |
| - | - | 1 | - |
| PM6, PS4\_supporting | ClinVar | 2 | 2 |
| - | - | 1 | 0 |
| - | - | - | - |
| PM6 | ClinVar | 2 | 1 |
| PM6 | ClinVar | 2 | 1 |
|  | TOTAL | 137 | 50 |

|  |  |  |
| --- | --- | --- |
|  | Final classification |  |
| Not found | ClinGen | CardioClassifier |
| 0 | Likely Pathogenic | Likely Pathogenic |
| 2 | VUS | VUS |
| 1 | Pathogenic | Likely Pathogenic |
| 0 | Likely Pathogenic | Likely Pathogenic |
| 0 | Likely Pathogenic | Likely Pathogenic |
| 0 | Pathogenic | Pathogenic |
| 1 | Pathogenic | Pathogenic |
| 0 | Pathogenic | Pathogenic |
| 1 | Pathogenic | Likely Pathogenic |
| 2 | VUS | VUS |
| 1 | Pathogenic | Pathogenic |
| 0 | Pathogenic | Pathogenic |
| 0 | Pathogenic | Pathogenic |
| 0 | Pathogenic | Pathogenic |
| 0 | Pathogenic | Pathogenic |
| 1 | Pathogenic | Pathogenic |
| 1 | Benign | Benign |
| 1 | Pathogenic | Pathogenic |
| 0 | Pathogenic | Pathogenic |
| 0 | Likely Pathogenic | Likely Pathogenic |
| 1 | Pathogenic | Likely Pathogenic |
| 1 | Pathogenic | Pathogenic |
| 0 | VUS | VUS |
| 2 | Pathogenic | Likely Pathogenic |
| 1 | Likely Pathogenic | VUS |
| 1 | Likely Pathogenic | Likely Pathogenic |
| 0 | Likely Pathogenic | Likely Pathogenic |
| 0 | Pathogenic | Pathogenic |
| 2 | Likely Pathogenic | Likely Pathogenic |
| 0 | Pathogenic | Pathogenic |
| 0 | Pathogenic | Pathogenic |
| 0 | Pathogenic | Pathogenic |
| 0 | Likely Pathogenic | Likely Pathogenic |
| 1 | VUS | VUS |
| 1 | VUS | VUS |
| 1 | VUS | VUS |
| 0 | VUS | VUS |
| 0 | VUS | VUS |
| 0 | VUS | VUS |
| 1 | Likely Pathogenic | VUS |
| 0 | VUS | VUS |
| 2 | Likely Pathogenic | Likely Pathogenic |
| 1 | Likely Pathogenic | Likely Pathogenic |
| 0 | VUS | VUS |
| 1 | VUS | VUS |
| 1 | VUS | VUS |
| 1 | VUS | VUS |
| 0 | VUS | VUS |
| 1 | Likely Pathogenic | Likely Pathogenic |
| 0 | VUS | VUS |
| 1 | VUS | VUS |
| 1 | VUS | VUS |
| 0 | Likely Pathogenic | Likely Pathogenic |
| 2 | VUS | VUS |
| 0 | VUS | VUS |
| 0 | VUS | Likely Pathogenic |
| 1 | Likely Pathogenic | Likely Pathogenic |
| 35 |  |  |

**Supplementary Table S2: Full results from comparison of CardioClassifier to ClinGen MYH7 pilot variants.** Rules were split into those that can be computationally annotated and those that are 'case-level' and require manual input. CardioClassiifer was ran using an 'All Cardiomyopathy' test to reflect the spectrum of phenotypes caused by variants in MYH7. Automatic retrieval of computational rules by CardioClassifier was assessed along with retrieval of case-level data through links present on the report output. \*Of the computational rules, PVS1\_moderate and BS4 were removed from the analysis as we would not expect these to be retrieved by CardioClassifier.

Table continues below

|  |  |  |  |
| --- | --- | --- | --- |
| Disease | Number of Variants | CardioClassifier Path/LikelyPath | CardioClassifier Proportion |
| HCM | 158 | 121 | 0.766 |
| DCM\* | 16 | 7 | 0.438 |
| ARVC | 22 | 17 | 0.773 |
| LQTS | 18 | 8 | 0.444 |
| CPVT | 1 | 1 | 1 |
| BrS | 4 | 2 | 0.5 |
| TOTAL | 219 | 156 | 0.712 |

|  |  |
| --- | --- |
| InterVar Path/LikelyPath | InterVar Proportion |
| 52 | 0.329 |
| 5 | 0.313 |
| 6 | 0.273 |
| 0 | 0 |
| 0 | 0 |
| 1 | 0.25 |
| 64 | 0.292 |

**Supplementary Table S3: Full results from comparison of CardioClassifier to InterVar.** Comparison was performed for a set of 219 variants identified as ‘Pathogenic’ or ‘Likely Pathogenic’ in ClinVar. \*Variant numbers for DCM are low, this is because mutations in TTN, the most common cause of DCM, are most often unique to an individual patient and their family. Our stringent selection criteria, which requires multiple pathogenic reports of the same variants, does therefore not include these.

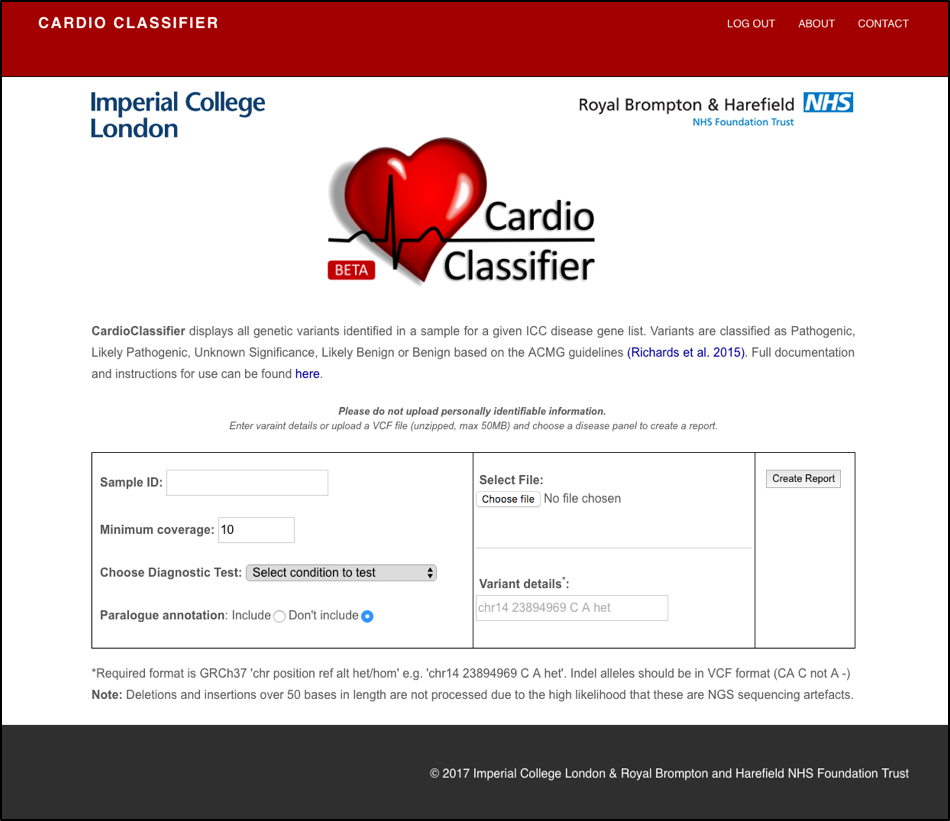
Table continues below

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Chr | Position | Ref | Alt | Gene | HGVS | Disease | Rule |
| 11 | 47364249 | G | A | MYBPC3 | c.1504C>T | HCM | PP1 |
| 11 | 47369975 | C | T | MYBPC3 | c.772G>A | HCM | PS3 |
| 11 | 47369975 | C | T | MYBPC3 | c.772G>A | HCM | PP1 |
| 11 | 47359280 | C | CC | MYBPC3 | c.2373dupG | HCM | PS3 |
| 11 | 47359280 | C | CC | MYBPC3 | c.2373dupG | HCM | PP1 |
| 11 | 47364129 | C | G | MYBPC3 | c.1624G>C | HCM | PS3 |
| 11 | 47364129 | C | G | MYBPC3 | c.1624G>C | HCM | PP1 |
| 11 | 47361343 | T | C | MYBPC3 | c.1928-2A>G | HCM | PS3 |
| 11 | 47361343 | T | C | MYBPC3 | c.1928-2A>G | HCM | PP1 |
| 11 | 47359115 | C | T | MYBPC3 | c.2429G>A | HCM | PP1 |
| 11 | 47360926 | TG | T | MYBPC3 | c.2096delC | HCM | PP1 |
| 11 | 47356671 | G | A | MYBPC3 | c.2827C>T | HCM | PP1 |
| 11 | 47354848 | T | TA | MYBPC3 | c.3226\_3227insT | HCM | PP1 |
| 11 | 47356632 | CAG | C | MYBPC3 | c.2864\_2865delCT | HCM | PS3 |
| 11 | 47356632 | CAG | C | MYBPC3 | c.2864\_2865delCT | HCM | PP1 |
| 11 | 47364269 | C | T | MYBPC3 | c.1484G>A | HCM | PP1 |
| 11 | 47353740 | G | A | MYBPC3 | c.3697C>T | HCM | PP1 |
| 11 | 47354743 | A | C | MYBPC3 | c.3330+2T>G | HCM | PS3 |
| 11 | 47367923 | T | C | MYBPC3 | c.927-2A>G | HCM | PS3 |
| 11 | 47367923 | T | C | MYBPC3 | c.927-2A>G | HCM | PP1 |
| 11 | 47354116 | C | T | MYBPC3 | c.3627+1G>A | HCM | PS3 |
| 11 | 47354116 | C | T | MYBPC3 | c.3627+1G>A | HCM | PP1 |
| 11 | 47364248 | C | T | MYBPC3 | c.1505G>A | HCM | PP1 |
| 11 | 47356592 | C | T | MYBPC3 | c.2905+1G>A | HCM | PS3 |
| 11 | 47357495 | C | T | MYBPC3 | c.2670G>A | HCM | PP1 |
| 11 | 47359085 | C | T | MYBPC3 | c.2459G>A | HCM | PP1 |
| 11 | 47370037 | T | G | MYBPC3 | c.710A>C | HCM | PP1 |
| 14 | 23894525 | C | T | MYH7 | c.2389G>A | HCM | PP1 |
| 14 | 23896866 | C | T | MYH7 | c.1816G>A | HCM | PP1 |
| 14 | 23896866 | C | T | MYH7 | c.1816G>A | HCM | PM6 |
| 14 | 23899059 | C | T | MYH7 | c.1063G>A | HCM | PP1 |
| 14 | 23893268 | C | T | MYH7 | c.2770G>A | HCM | PS3 |
| 14 | 23893268 | C | T | MYH7 | c.2770G>A | HCM | PP1 |
| 14 | 23893268 | C | T | MYH7 | c.2770G>A | HCM | PS2 |
| 14 | 23900811 | C | T | MYH7 | c.715G>A | HCM | PP1 |
| 14 | 23887453 | C | T | MYH7 | c.4135G>A | HCM | PP1 |
| 14 | 23894612 | C | T | MYH7 | c.2302G>A | HCM | PP1 |
| 19 | 55665514 | G | A | TNNI3 | c.433C>T | HCM | PS3 |
| 19 | 55665514 | G | A | TNNI3 | c.433C>T | HCM | PP1 |
| 19 | 55665462 | C | T | TNNI3 | c.485G>A | HCM | PP1 |
| 19 | 55665477 | G | A | TNNI3 | c.470C>T | HCM | PP1 |
| 19 | 55665525 | C | T | TNNI3 | c.422G>A | HCM | PP1 |
| 1 | 201332504 | TCTC | T | TNNT2 | c.487\_489delGAG | HCM | PP1 |
| 1 | 201334766 | A | T | TNNT2 | c.236T>A | HCM | PP1 |
| 1 | 201334426 | G | A | TNNT2 | c.274C>T | HCM | PP1 |
| 15 | 63353922 | G | A | TPM1 | c.574G>A | HCM | PP1 |
| 15 | 63353098 | G | A | TPM1 | c.523G>A | HCM | PS3 |
| 15 | 63353098 | G | A | TPM1 | c.523G>A | HCM | PP1 |
| 15 | 63353098 | G | A | TPM1 | c.523G>A | HCM | PM6 |
| 1 | 201331098 | ATCT | A | TNNT2 | c.629\_631delAGA | DCM | PP1 |
| 1 | 201333464 | G | A | TNNT2 | c.421C>T | DCM | PS3 |
| 1 | 201333464 | G | A | TNNT2 | c.421C>T | DCM | PP1 |
| 1 | 201333464 | G | A | TNNT2 | c.421C>T | DCM | PM6 |
| 12 | 32955491 | C | G | PKP2 | c.2146-1G>C | ARVD/C | PS3 |
| 12 | 32955491 | C | G | PKP2 | c.2146-1G>C | ARVD/C | PP1 |
| 12 | 32955434 | GGTGTG | C | PKP2 | c.2197\_2202delCACACCinsG | ARVD/C | PP1 |
| 12 | 32949042 | C | T | PKP2 | c.2489+1G>A | ARVD/C | PP1 |

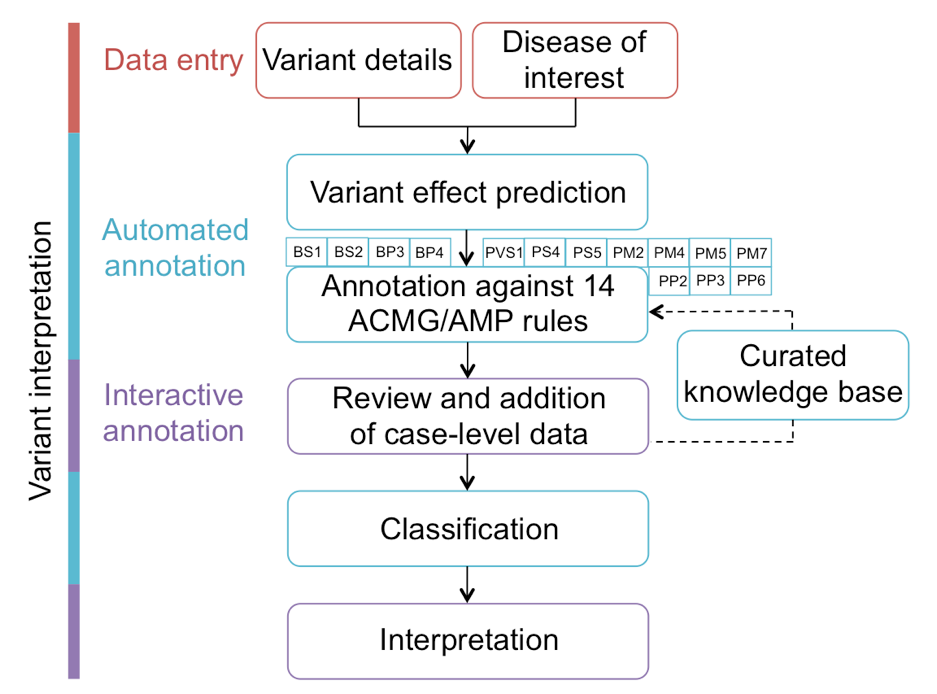
|  |  |
| --- | --- |
| Strength | Source |
| Strong | <PMID:20378854> |
| Strong | <PMID:19590044,15114369,19574547,25031304> |
| Strong | <PMID:9562578,16858239> |
| Strong | <PMID:19273718,27108529> |
| Strong | ClinVar:SCV000059136.4;<PMID:10736283,9562578> |
| Strong | <PMID:9048664,10610770,22057632> |
| Strong | ClinVar:SCV000256661.2,SCV000059072.4;<PMID:9048664> |
| Strong | <PMID:10610770,11499719,7493026> |
| Strong | ClinVar:SCV000059101.4 |
| Supporting | ClinVar:SCV000256161.1 |
| Moderate | ClinVar:SCV000059112.4 |
| Supporting | ClinVar:SCV000059178.4 |
| Moderate | ClinVar:SCV000059217.4 |
| Strong | <PMID:19273718> |
| Strong | ClinVar:SCV000256180.1 |
| Supporting | ClinVar:SCV000198912.3 |
| Strong | <PMID:11499719,16199542,18957093,16715312> |
| Strong | <PMID:17937428> |
| Strong | <PMID:9562578> |
| Strong | ClinVar:SCV000059333.4 |
| Strong | <PMID:9562578> |
| Strong | <PMID:9562578> |
| Strong | <PMID:9562578> |
| Strong | <PMID:11499719> |
| Supporting | ClinVar:SCV000059169.4 |
| Moderate | ClinVar:SCV000198854.3 |
| Strong | ClinVar:SCV000059314.4 |
| Strong | ClinVar:SCV000059441.4 |
| Strong | ClinVar:SCV000280305.1 |
| Moderate | ClinVar:SCV000208754.6 |
| Supporting | ClinVar:SCV000208717.9 |
| Strong | <PMID:10024460> |
| Supporting | ClinVar:SCV000212642.1 |
| Strong | <PMID:1430197,1552912> |
| Moderate | ClinVar:SCV000059647.4 |
| Strong | ClinVar:SCV000059538.4,SCV000212637.1 |
| Moderate | ClinVar:SCV000199202.3 |
| Strong | <PMID:19651143> |
| Strong | ClinVar:SCV000059946.4 |
| Strong | ClinVar:SCV000059952.4 |
| Strong | ClinVar:SCV000059951.4 |
| Supporting | ClinVar:SCV000209166.6 |
| Strong | ClinVar:SCV000060251.4 |
| Strong | <PMID:7898523,18651846> |
| Strong | ClinVar:SCV000060228.4 |
| Strong | ClinVar:SCV000060282.4;<PMID:26960954> |
| Strong | <PMID:9440709,10400910,10900175> |
| Strong | ClinVar:SCV000059992.4 |
| Moderate | <PMID:7729014> |
| Strong | ClinVar:SCV000060262.4 |
| Strong | <PMID:18349139,18606313> |
| Strong | <PMID:11684629,15769782> |
| Moderate | <PMID:26656454> |
| Strong | <PMID:15489853,19279339,25087486> |
| Moderate | <PMID:17010805,25087486,27335691> |
| Moderate | ClinVar:SCV000061871.4 |
| Moderate | ClinVar:SCV000199907.3 |

**Supplementary Table S4: Manually curated evidence added to the CardioClassifier knowledge-base.** For 63 variants from the ACGV resource that had not been curated as part of the ClinGen MYH7 pilot we searched the literature and ClinVar for evidence of segregation, *de novo* occurrence or functional evidence. This evidence is uploaded into CardioClassifier so that it appears automatically when one of these variants in searched.

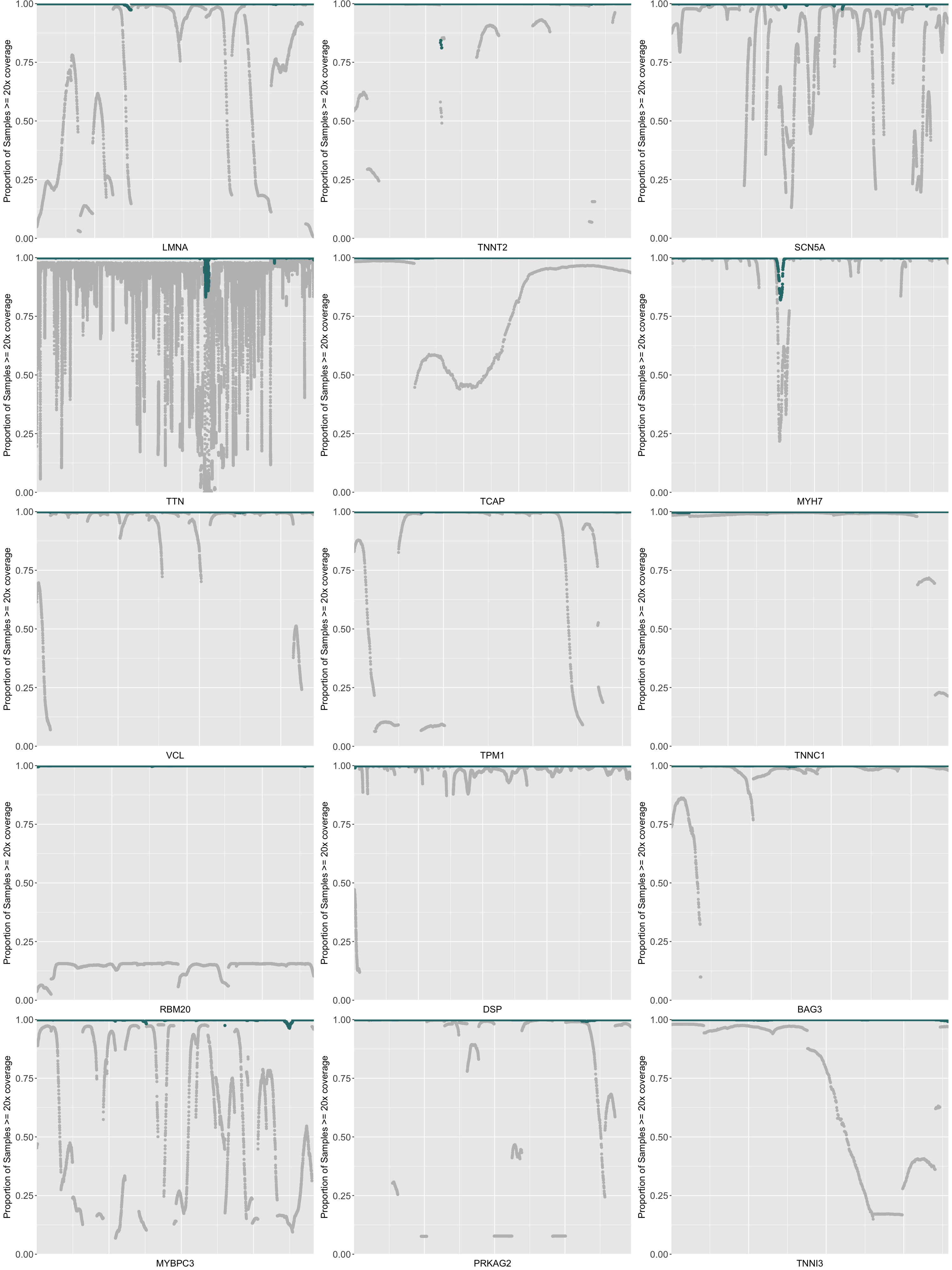
## SUPPLEMENTARY FIGURES

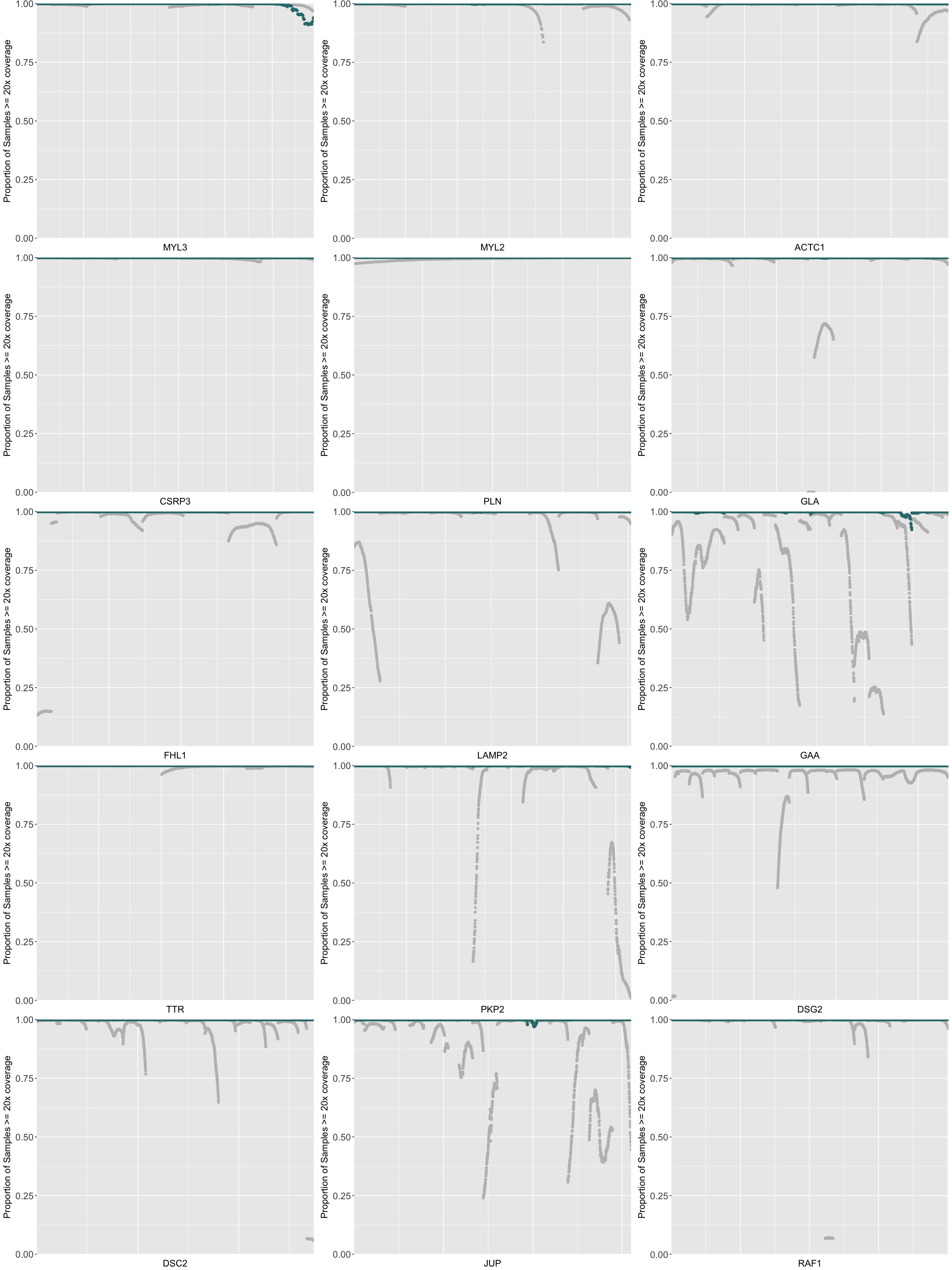


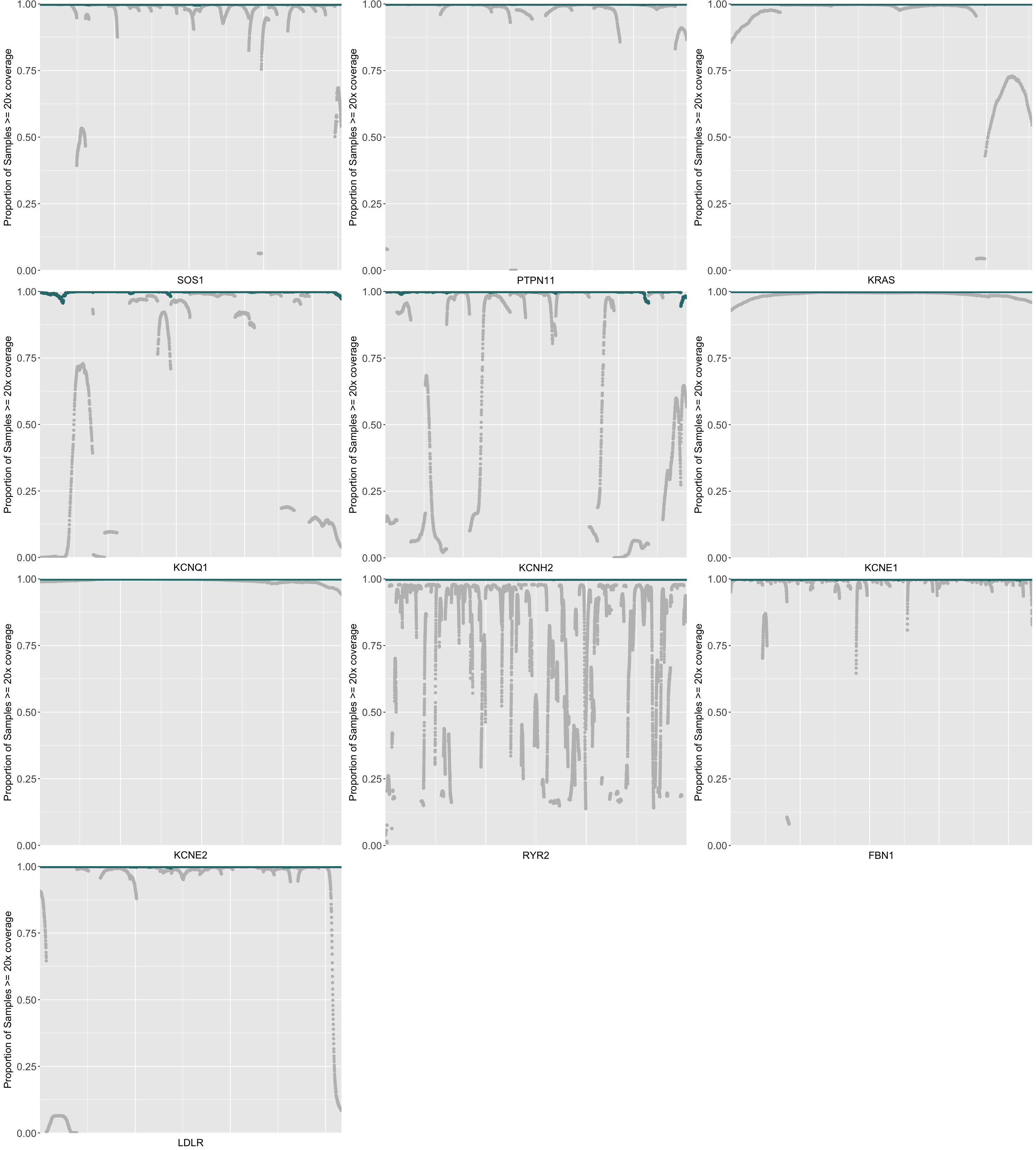
**Supplementary Figure S1: The CardioClassifier web-interface.** Users input single variant details or upload a variant call file (VCF) for a sample and select a disease or test of interest.



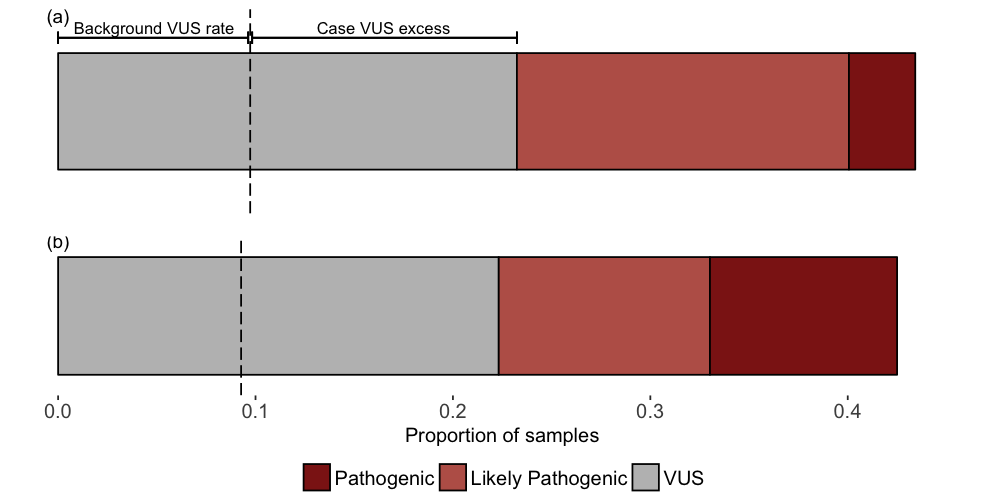
**Supplementary Figure S2: Flow diagram showing the process of variant interpretation using CardioClassifier.** Users input variant details and select a disease of interest before variants are annotated by Ensembl’s Variant Effect Predictor (VEP) and against 17 computational criteria.







**Supplementary Figure S3: Per gene comparison of coverage in ExAC and the TruSight Cardio sequencing panel.** Plotted are the coding regions of each gene from lowest genomic position to highest (i.e. not considering strand) against the proportion of samples with >20x coverage.



**Supplementary Figure S4: Potentially pathogenic variants from 327 HCM cases.** The figure shows the proportion of cases with Pathogenic and Likely Pathogenic variants and variants of uncertain significance (VUS) in HCM cases called by CardioClassifier before (a) and after (b) addition of curated case-level and functional data. The dotted line represents the background rate of variation for a HCM test i.e. the proportion of 625 healthy volunteer samples with a VUS.