Clinical validation of large-scale functional assays: CanVIG-UK recommendations for truthset assembly based on 2122 gene-truthset-assay evaluations

Sophie Allen[1], Charlie F. Rowlands[1], Alice Garrett[1,2], Miranda Durkie[3], George J. Burghel[4,5], Rachel Robinson[6], Alison Callaway[7], Joanne Field[8], Bethan Frugtniet[2], Sheila Palmer-Smith[9], Jonathan Grant[10], Judith Pagan[11], Elizabeth Johnston[8], Trudi McDevitt[12], Lowri Hughes[13], Laura Yarram-Smith[14], Peter Logan[15], Laura Reed[16], Katie Snape[2], Terri McVeigh[17], Helen Hanson[18,19], Frederick P. Roth[20], Lea M. Starita[21,22], Rehan Villani[23], Amanda B. Spurdle[23,24], David J. Adams[25], Greg Findlay[26], Clare Turnbull[1,17] and CanVIG-UK

Introduction

- Evidence from functional assays can be used towards a clinical variant classification using the ACMG/AMP 2015 standards (1)
- The strength of evidence applicable is influenced by the **number and type** of available **pathogenic and benign** variants – so-called '**truthset**' or '**reference**' variants – per the Brnich et al 2019 guidelines (2)
- '**Truthsets**' can be constructed in many different ways
- A **lack of clarity, transparency and consistency** in the stipulations and standards for assembly of truthsets therefore represents a **major hurdle for clinical diagnostic laboratories** to use new MAVE data
- We evaluated >2000 truthset-gene-assay combinations to **explore the variance in evidence strength** using difference permutations of truthsets

Methods

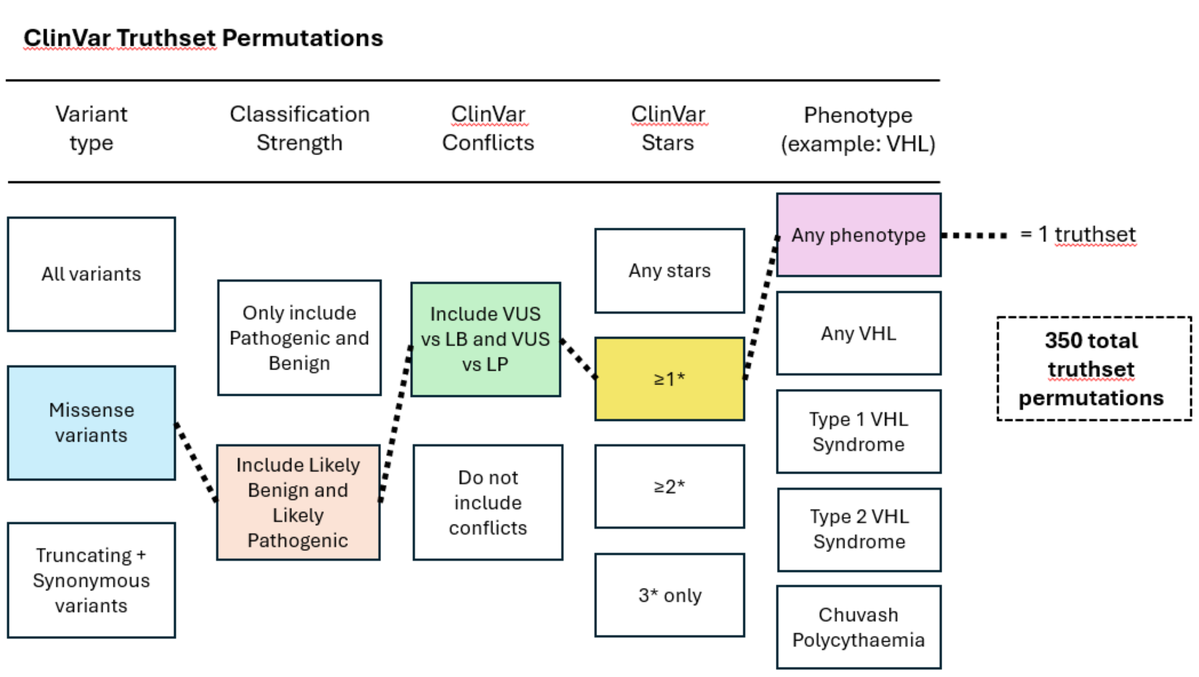

Figure 1: Flowchart of ClinVar permutation options

- For four cancer susceptibility genes (*BRCA1, BRCA2, RAD51C*, and *VHL*), we defined 70 truthset permutations using ClinVar. (Figure 1)
- We identified additional 'proxy-benign' missense variants using a combination of population data, case-control evidence, and predictive tools (3), and added these to ClinVar pathogenic variants to create additional truthsets.
- For *VHL*, we also explored truthsets submitted to ClinVar against five phenotypes, defining 350 truthset permutations.
- We used these truthsets to calculate evidence strength for five large-scale functional assays following the Brnich et al methodology (2).

Results

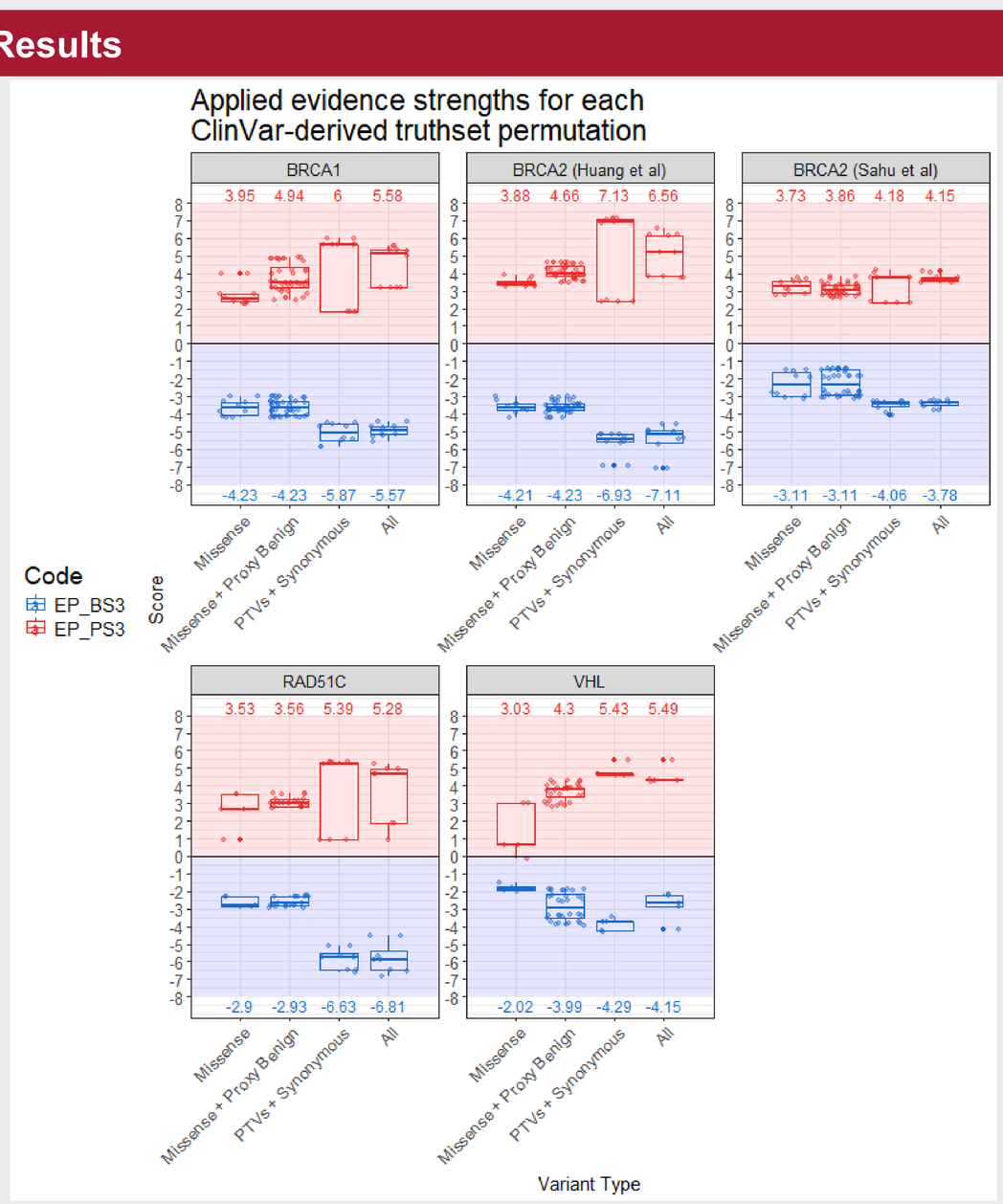

Figure 2: Boxplots of evidence points assigned for each ClinVar truthset permutation for each assay examined. Truthsets are separated by variant type included, where "Missense+proxy-benign" includes systematically generated benign missense truthsets. For each truthset group, the maximum evidence towards pathogenicity (PS3) is displayed at the top of the plot, and the maximum evidence towards benignity is displayed at the bottom.

- Truthsets containing protein truncating variants (PTVs) and synonymous variants attained stronger evidence overall
- Higher stringency truthsets (eg 2* ClinVar vs 1* ClinVar classifications) resulted in fewer false positive/false negative readouts, but had fewer variants and lower power, thus evidence strength was not increased.
- Inclusion of proxy-benign missense variants increased maximum evidence strength available for all assays examined.
- We also reviewed CSPEC pages for 112 genes across 39 VCEPs – there is high variability in how functional data is applied and in truthset construction:
  - Truthsets were provided for 44 guidelines (56%)
  - Truthset provenance was a mixture of VCEP classifications, cases, ClinVar classifications, variant databases, and variants from the assay publication.

Conclusion

Truthset construction impacts the evidence strength applicable for a functional assay, and the trade-off between power and stringency should be considered in future recommendations for truthset construction.

Contact

Sophie Allen, MSc | PhD Student
Division of Genetics and Epidemiology
The Institute of Cancer Research
32 Oakleaf Avenue, Sutton, London, UK
Sophie.Allen@icr.ac.uk

Author Affiliations

1 Division of Genetics and Epidemiology, The Institute of Cancer Research, London, UK; St George's University Hospitals NHS Foundation Trust, Tooting, London, UK; North East and Yorkshire Genomic Laboratory Hub, Sheffield Children's NHS Foundation Trust, Sheffield, UK; Division of Cancer Sciences, School of Medical Sciences, Faculty of Biology, Medicine and Health, The University of Manchester, Manchester, UK; Manchester Centre for Genomic Medicine and NW Laboratory Genetics Hub, Manchester University Hospitals NHS Foundation Trust, Manchester, UK; The Leeds Genetics Laboratory, NEY Genomic Laboratory Hub, Leeds Teaching Hospitals NHS Trust, Leeds, UK; Central and South Genomics Laboratory Hub, Wessex Genomics Laboratory Service, University Hospital Southampton NHS Foundation Trust, Salisbury, UK; Genomics and Molecular Medicine Service, Nottingham University Hospitals NHS Trust, Nottingham, UK; Wales Genomic Health Centre, Cardiff and Vale University Health Board, Cardiff, UK; Laboratory Genetics, Queen Elizabeth University Hospital, NHS Greater Glasgow and Clyde, Glasgow, UK; South East Scotland Clinical Genetics, Western General Hospital, Edinburgh, UK; Department of Clinical Genetics, CHI at Crumlin, Dublin, Ireland; West Midlands Genomics Laboratory, Birmingham Women's and Children's NHS Foundation Trust, Birmingham, UK; North Bristol NHS Trust, Southmead Hospital, Bristol, UK; Belfast Health and Social Care Trust, Royal Victoria Hospital, Belfast, UK; Rare & Inherited Disease Laboratory, NHS North Thames Genomic Laboratory Hub, Great Ormond Street Hospital for Children NHS Foundation Trust, London, UK; The Royal Marsden NHS Foundation Trust, Fulham Road, London, UK; Peninsula Regional Genetics Service, Royal Devon University Healthcare NHS Foundation Trust, Exeter, UK; Department of Clinical and Biomedical Sciences, University of Exeter Medical School, Exeter, UK; Department of Computational and Systems Biology, University of Pittsburgh School of Medicine, Pittsburgh, PA, USA; Department of Genome Sciences, University of Washington, Seattle, WA, USA; Brotman Baty Institute for Precision Medicine, Seattle, WA, USA; Population Health Program, QIMR Berghofer, Brisbane, QLD, Australia; The University of Queensland, Brisbane, QLD, Australia; Wellcome Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge, UK; The Genome Function Laboratory, The Francis Crick Institute, London, UK

References

(1) Richards et al., 2015. Genetics in Medicine. DOI: 10.1038/gim.2015.30
(2) Brnich et al., 2019. Genome Medicine. DOI: 10.1186/s13073-019-0690-2
(3) Rowlands et al., 2025. AJHG. DOI: 10.1016/j.ajhg.2025.08.014