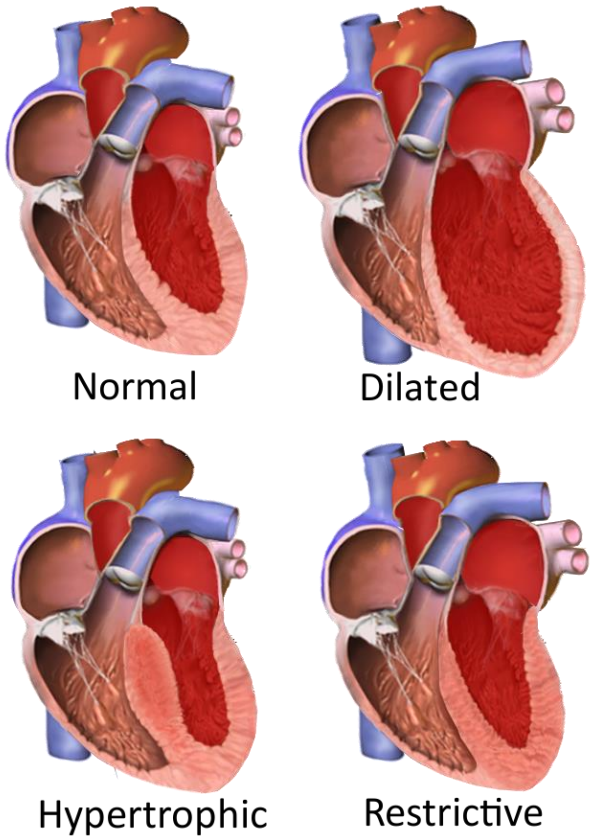
Assessing the impact of variants in Cardiomyopathies and the role of the Obscurin gene

TIMIR WESTON

Obscurin and Cardiomyopathy



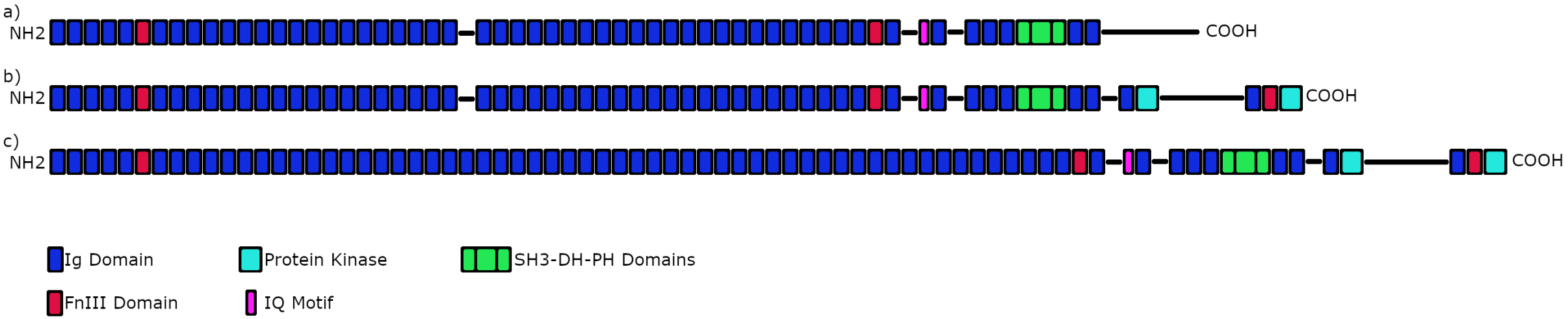
Via: *Blausen.com staff (2014). "Medical gallery of Blausen Medical 2014".* WikiJournal of Medicine ***1*** *(2).*

- Obscurins are a family of modular proteins expressed from *OBSCN* gene – 40-800kDa
- Damaging variants implicated in cardiomyopathies – diseases of the myocardium
- SNVs in giant proteins are hard to classify
  - "Rare" alleles can appear common
  - May only be damaging in tandem with others
  - Frequency differs between sub-populations
- For titin, a resource that enabled mapping of SNVs to sequence/structural information was created to alleviate some of these issues - TITINdb

Overview

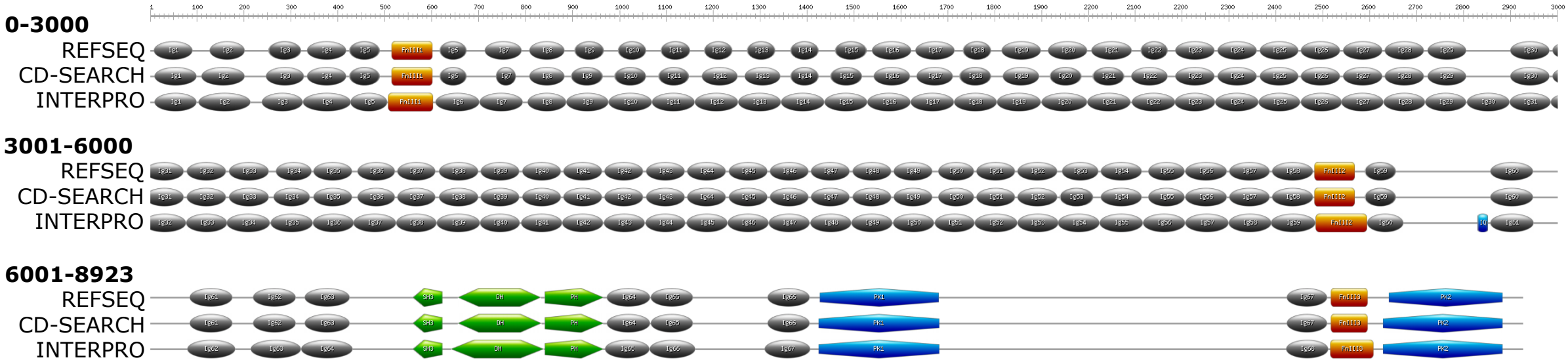
- Establishing the Domain Architecture
- Properties of the Resource
- Potential Applications: Methods
- Potential Applications: Results
  - Rare vs Common SNVs
  - Protein stability change and hotspots

Establishing the Domain Architecture



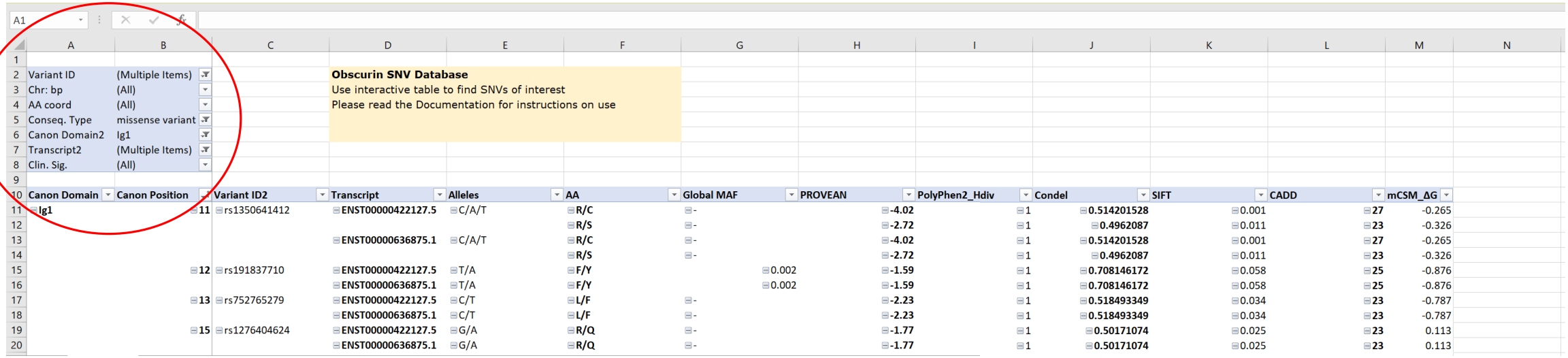
a) Obscurin-a (Isoform 3)
b) Obscurin-b (Isoform 1)
c) Obscurin-IC (inferred complete)

Establishing the Domain Architecture



- Current lack of consensus over domain boundaries
- We used analysis programs PHMMER, NCBI CD-Search, InterPro 72.0 to construct domain architectures for obscurin-IC
- Architectures compared with bioinformatics resources, experimental structures, each other
- Final architecture used CD-Search domains as a base (highest alignment score for Ig domains)

A centralised resource for obscurin SNVs



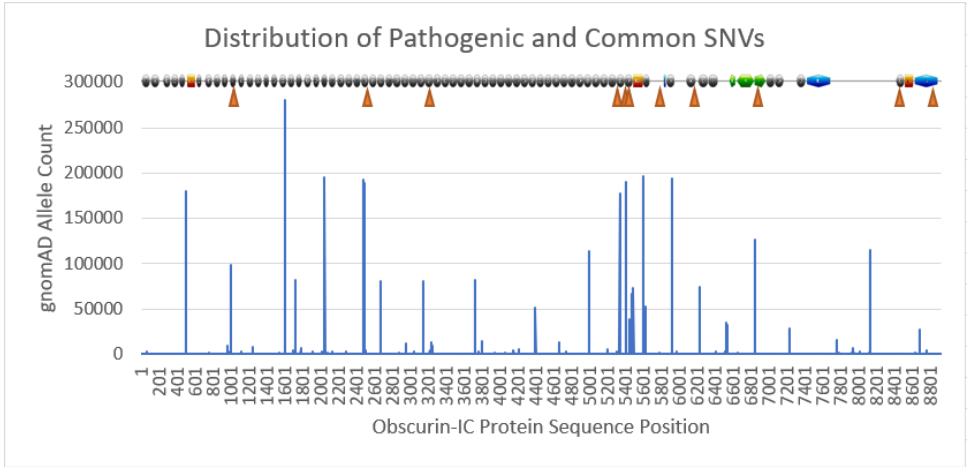
	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1														
2	Variant ID	(Multiple Items)		**Obscurin SNV Database**										
3	Chr: bp	(All)		Use interactive table to find SNVs of interest										
4	AA coord	(All)		Please read the Documentation for instructions on use										
5	Conseq. Type	missense variant												
6	Canon Domain2	Ig1												
7	Transcript2	(Multiple Items)												
8	Clin. Sig.	(All)												
9														
10	**Canon Domain**	**Canon Position**	**Variant ID2**	**Transcript**	**Alleles**	**AA**	**Global MAF**	**PROVEAN**	**PolyPhen2_Hdiv**	**Condel**	**SIFT**	**CADD**	**mCSM_ΔG**	
11	**Ig1**	**11**	rs1350641412	**ENST00000422127.5**	C/A/T	**R/C**	-	**-4.02**	1	**0.514201528**	0.001	**27**	-0.265	
12						**R/S**	-	**-2.72**	1	**0.4962087**	0.011	**23**	-0.326	
13				**ENST00000636875.1**	C/A/T	**R/C**	-	**-4.02**	1	**0.514201528**	0.001	**27**	-0.265	
14						**R/S**	-	**-2.72**	1	**0.4962087**	0.011	**23**	-0.326	
15		**12**	rs191837710	**ENST00000422127.5**	T/A	**F/Y**	0.002	**-1.59**	1	**0.708146172**	0.058	**25**	-0.876	
16				**ENST00000636875.1**	T/A	**F/Y**	0.002	**-1.59**	1	**0.708146172**	0.058	**25**	-0.876	
17		**13**	rs752765279	**ENST00000422127.5**	C/T	**L/F**	-	**-2.23**	1	**0.518493349**	0.034	**23**	-0.787	
18				**ENST00000636875.1**	C/T	**L/F**	-	**-2.23**	1	**0.518493349**	0.034	**23**	-0.787	
19		**15**	rs1276404624	**ENST00000422127.5**	G/A	**R/Q**	-	**-1.77**	1	**0.50171074**	0.025	**23**	0.113	
20				**ENST00000636875.1**	G/A	**R/Q**	-	**-1.77**	1	**0.50171074**	0.025	**23**	0.113	

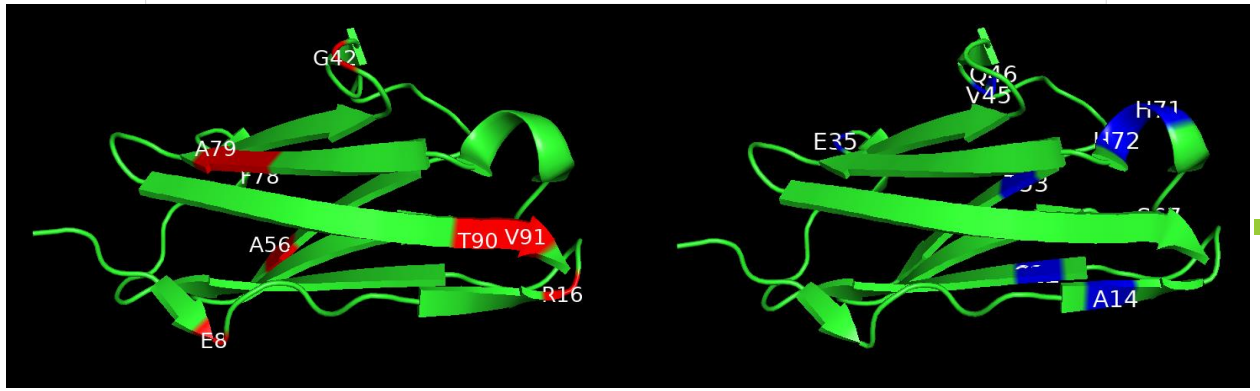
- Sources
  - Ensembl, gnomAD
- Pathogenicity Metrics
  - SIFT, PolyPhen-2, CADD, PROVEAN, Condel
  - mCSM, DUET
- Homology Modelling
  - BLAST, T-Coffee, SWISS-MODEL
- Implementation
  - Simple Excel functionality
  - Variant ID, position, consequence, MAF, pathogenicity metrics

Potential Applications: SNV sequence and structural distribution

- Resource should improve variant prioritisation for disease association
  - Integration of sequence and structural information
  - Disease-associated variant hotspots
- To illustrate the future uses that this resource could be put to, we investigated the following:
  - Comparison of pathogenic and common SNV locations on protein sequence
  - Comparison of pathogenic and common SNV locations on domain structure
  - Effects of SNVs on protein stability within a domain structure, and the association with solvent accessibility

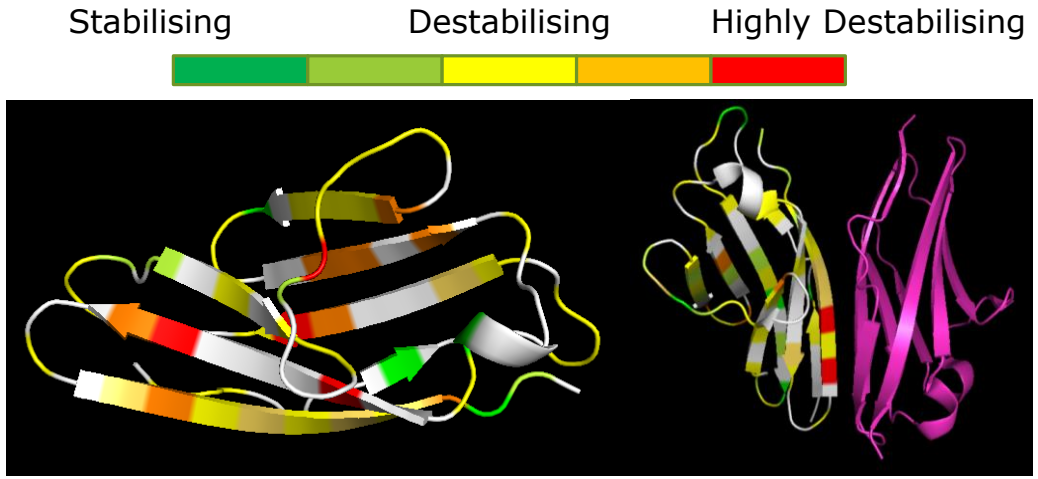
Pathogenic and Common SNVs tend to cluster in different domain locations

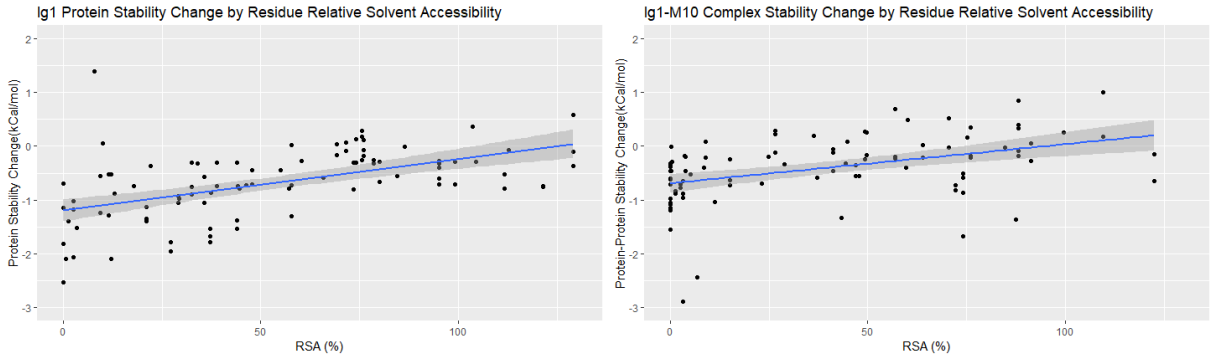




- Pathogenic and common SNVs do not seem to cluster in any shared locations, so pathogenic hotspots may exist in regions vital to protein stability
- The Ig58-Ig61 region has a clustering of both common and pathogenic, but the pathogenicity of two of these has been called into question
- Structural positions of Ig-domain SNVs also tend to be different for pathogenic and common SNVs

Protein stability change on mutation can be associated with domain location





- Protein stability change on mutation computed using mCSM
- Most destabilising regions differ for Ig1 alone and in complex with titin domain M10
- mCSM also computes relative solvent accessibility (RSA) for each residue
  - Linear regression gives minor correlation between stability change and RSA
  - Ig1: $R^2 = 0.285$, $p < 0.001$
  - Ig1-M10: $R^2 = 0.185$, $p < 0.001$
  - Violates homogeneity assumption

Conclusion

- We have created and implemented a domain architecture which allows for mapping of variants to structures with confidence
- Future work, as well as identification of pathogenic hotspots, should extend to the characterisation of hotspots for individual cardiomyopathies and dynamical predictors
- The locations of where binding partners of obscurin interact should be a focus of characterisation efforts, as they are integral to the structure of the sarcomere
- Improvements for the resource: modelling all domains in the sequence, in-built support for visualisation of SNVs on domain structure, addition of population and sub-population frequency data, ability to search SNVs by isoform, migration to web server for ease of access

Acknowledgements

- Professor Franca Fraternali
- Professor Mathias Gautel
- Fraternali Group – Joseph Ng, Irene Marzuoli
- BRC
- NIHR