

IDdb: A Database of Intellectual Disability Genes



Dias KR^{1,2}, Zhu Y³, Buckley M³, Evans C¹, Field M⁴, Roscioli T^{1,2,3,5}



¹ Neuroscience Research Australia, Sydney NSW, Australia ² Prince of Wales Clinical School, UNSW Medicine, Sydney NSW, Australia
³ SEALS Genetics, Prince of Wales Hospital, Sydney NSW, Australia ⁴ GoLD Service, Hunter Genetics, Newcastle NSW, Australia
⁵ Clinical Genetics, Sydney Children's Hospital, Sydney NSW, Australia

Background and Objectives

Every two hours, an Australian child will be diagnosed with intellectual disability¹.

Intellectual disability (ID):

- is prevalent in 1-3% of the population
- has early onset with frequent comorbidities
- is characterised by an IQ of <70

ID's extreme genetic and phenotypic heterogeneity leads to:

- a long diagnostic odyssey
- significant cost to families
- few specific diagnoses
- even fewer specific treatments
- economic cost of ~ \$14,720 billion annually in Australia²

Next generation sequencing technologies have:

- increased the diagnostic rate to 40-60%
- been able to sequence over 2000 genes have been associated with ID
- generates 1000's of variants that need to be manually curated

Access to a central repository of ID information would ensure this curation process was accurate and effective, but none exist.

IDdb aims to be a current, comprehensively-annotated catalogue of ID genes

Here we show how IDdb could be used to enable rapid variant curation, especially in prenatal and reproductive risk settings and highlight potential patterns in the underlying biology of ID, thereby catalysing the virtuous cycle between discovery and diagnosis in ID.

Methods

The literature curation and gene validation processes are based on the National Institute of Health's ClinGen framework and gene-disease validity standard operating procedure³.

Figure 1: Gene curation workflow

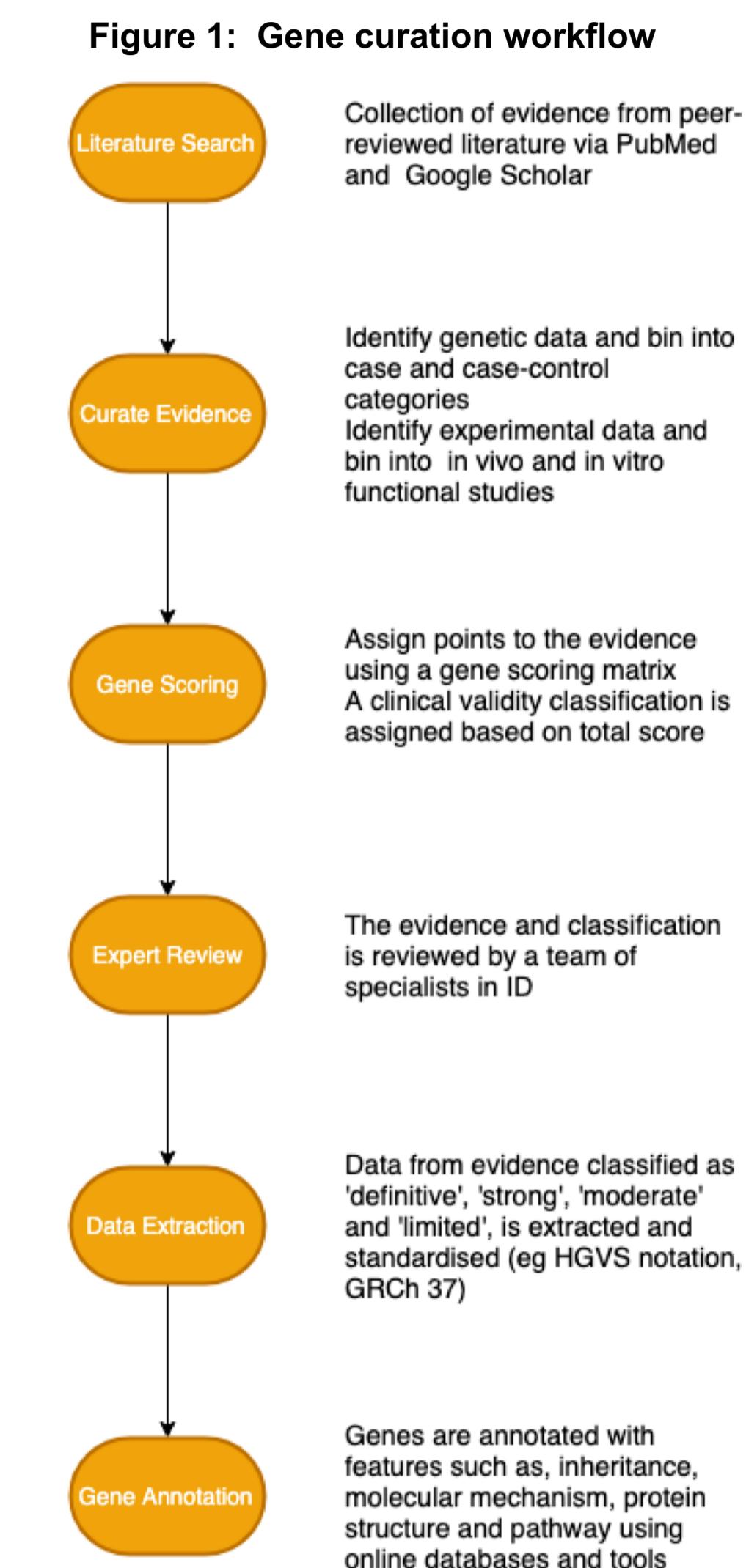


Table 1: Genetic evidence matrix

GENETIC EVIDENCE SUMMARY							
Evidence Type	Case Information (type of variant identified in proband)		Suggested Points/CASE		Points Given	Max Score	
	Default	Range					
	Variant is de novo	C 2	0-3	H	M 12		
Autosomal Dominant OR X-Linked Disorder A	Predicted or proven null variant	D 1.5	0-2	I	N 10		
	Other variant type (not predicted/proven null) with some evidence of gene impact	E 0.5	0-1.5	J	O 7		
Autosomal Recessive Disorder B	Two variants in trans and at least one de novo or predicted/proven null variant	F 2	0-3	K	P 12		
	Two variants (not predicted/proven null) with some evidence of gene impact in trans	G 1	0-1.5	L			
Segregation Evidence	Exome/Genome or all genes sequenced in linkage region	Sequencing Method	Q	R 3			
	Total LOD Score	Candidate Gene Sequencing					
	2-2.99	0.5					
Case-Control Data	3-4.99	1	S	T 12			
	>5	1.5					
	Evidence of segregation in one or more families						
Case-Control Study Type		Case-Control Quality Criteria		Suggested Points/Study	Points Given	Max Score	
Single Variant Analysis		<ul style="list-style-type: none"> Variant Detection Methodology Power Bias and Confounding Factors 		0-6	S	T 12	
Aggregate Variant Analysis		<ul style="list-style-type: none"> Statistical Significance 		0-6			
						TOTAL ALLOWABLE POINTS FOR GENETIC EVIDENCE U 12	

Table 2: Experimental evidence matrix

EXPERIMENTAL EVIDENCE SUMMARY						
Evidence Category	Evidence Type	Suggested Points/CASE		Points Given	Max Score	
		Default	Range			
Function	Biochemical Function	A 0.5	0-2	L	W 2	
	Protein Interaction	B 0.5	0-2			
	Expression	C 0.5	0-2			
Functional Alteration	Patient cells	D 1	0-2	N	X 2	
	Non-patient cells	E 0.5	0-1			
Models	Non-human model organism	F 2	0-4	Q	Y 4	
	Cell culture model	G 1	0-2			
Rescue	Rescue in human	H 2	0-4	T		
	Rescue in non-human model organism	I 2	0-4			
	Rescue in cell culture model	J 1	0-2			
						TOTAL ALLOWABLE POINTS FOR EXPERIMENTAL EVIDENCE Z 6

Results

IDdb consolidates data from multiple sources for gene annotation and feature information most relevant to ID. The literature search has resulted in over 2000 genes associated with ID that are currently being curated and assessed. The figures below show analysis derived from the top 100 ID genes, ranked on the sum of likely pathogenic and pathogenic CNVs and SNVs from the DECIPIER and ClinVar databases.

Figure 2: Features of IDdb

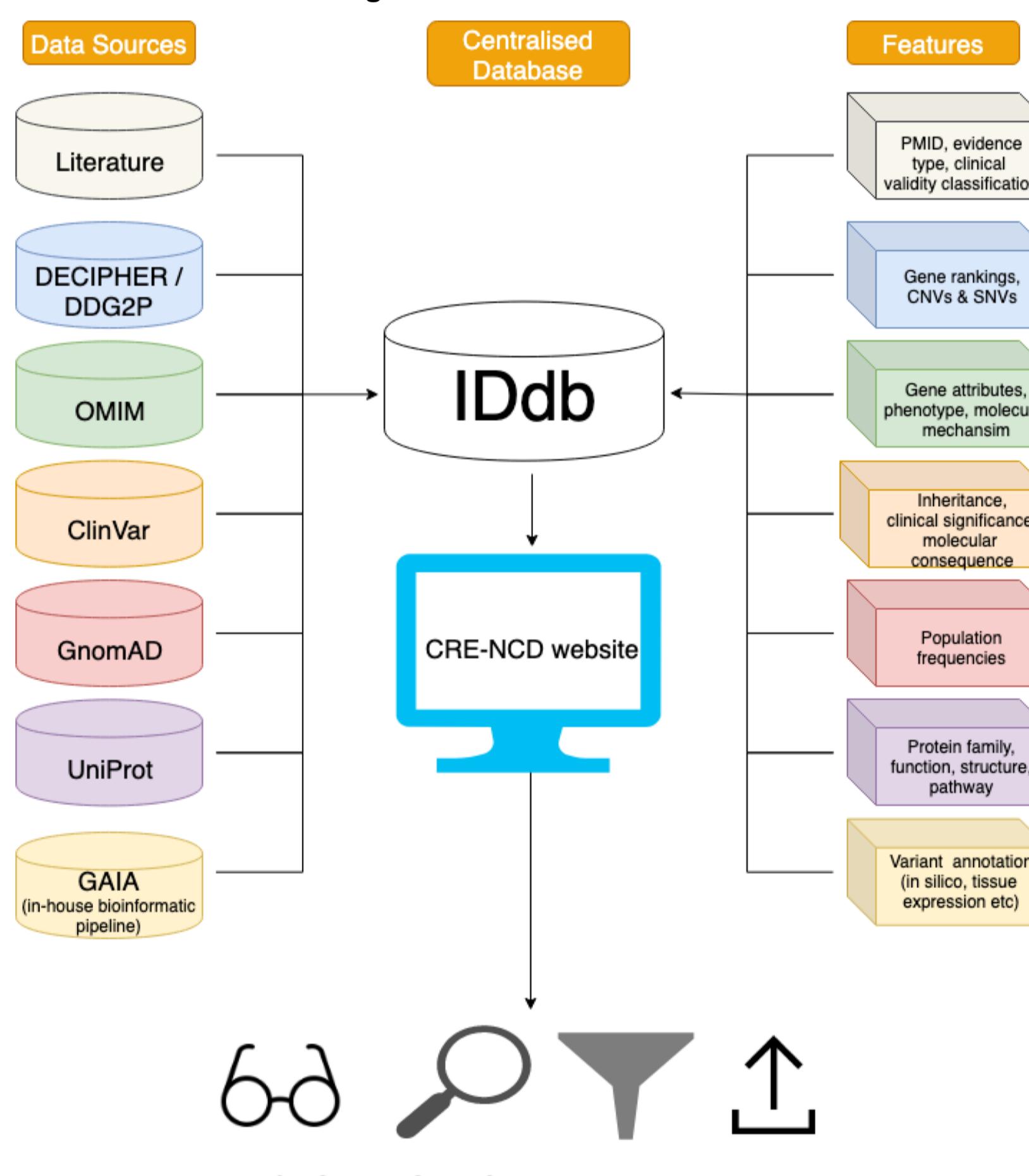


Figure 3: Number of pathogenic and likely pathogenic CNVs and SNVs in the top 100 ID genes

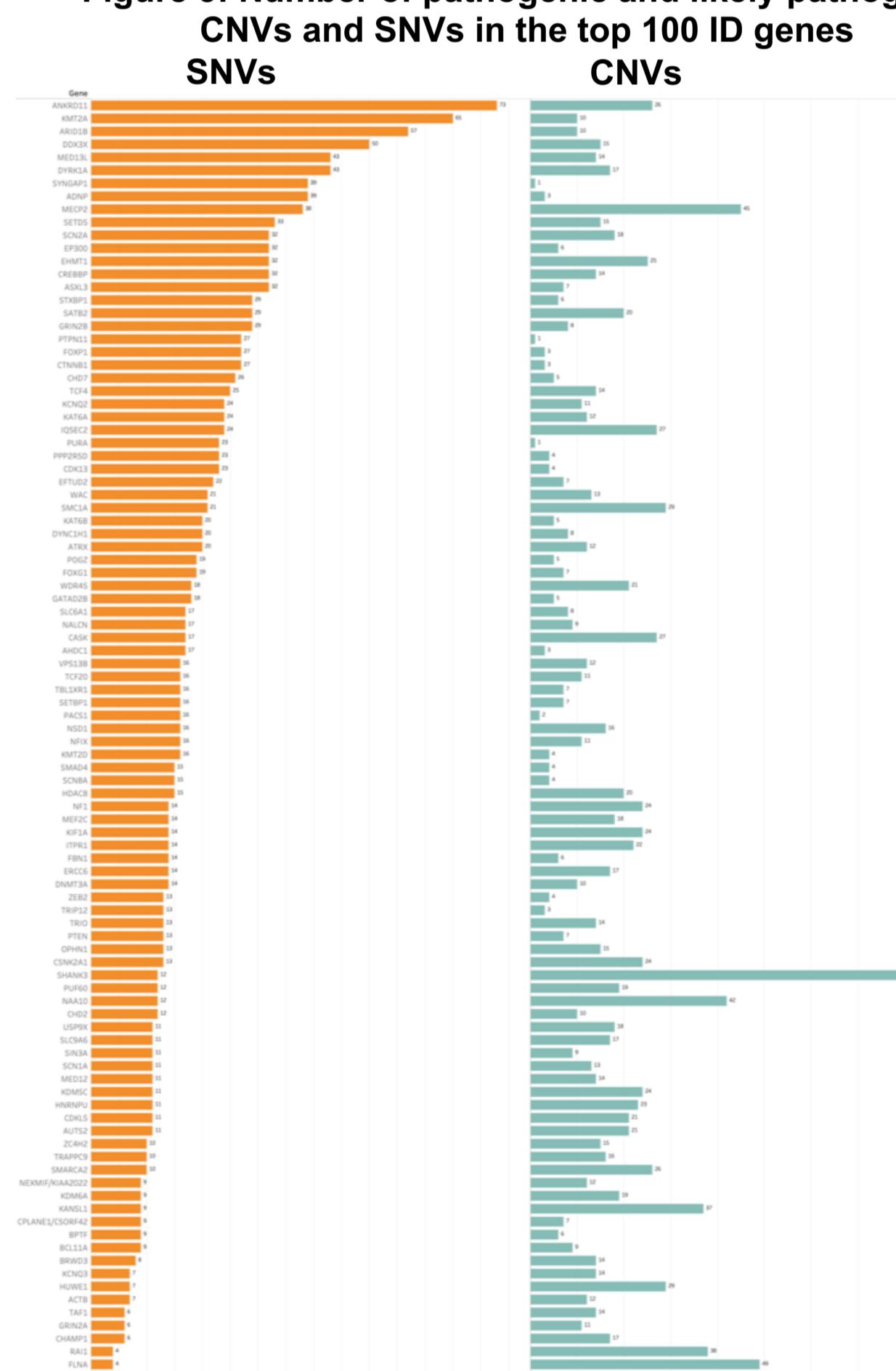


Figure 4: Number of top 100 ID genes by disease group

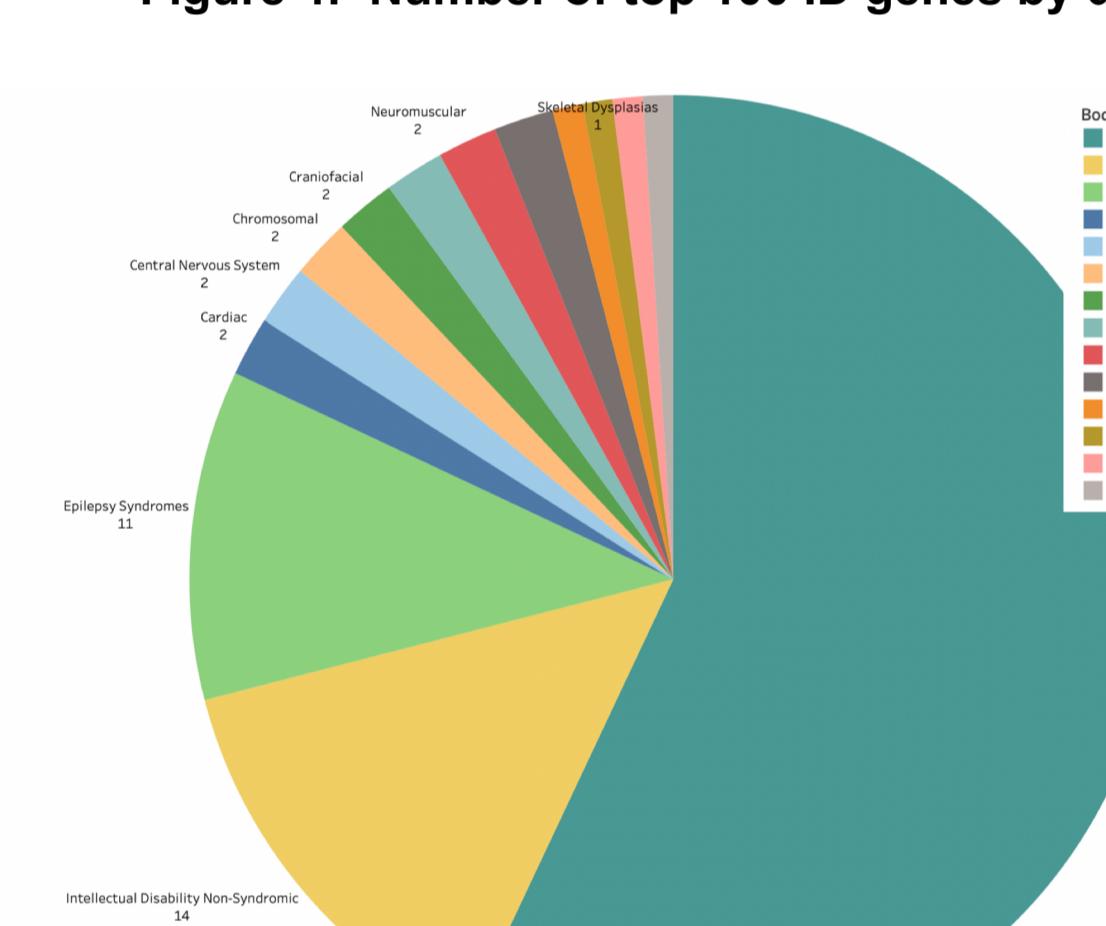
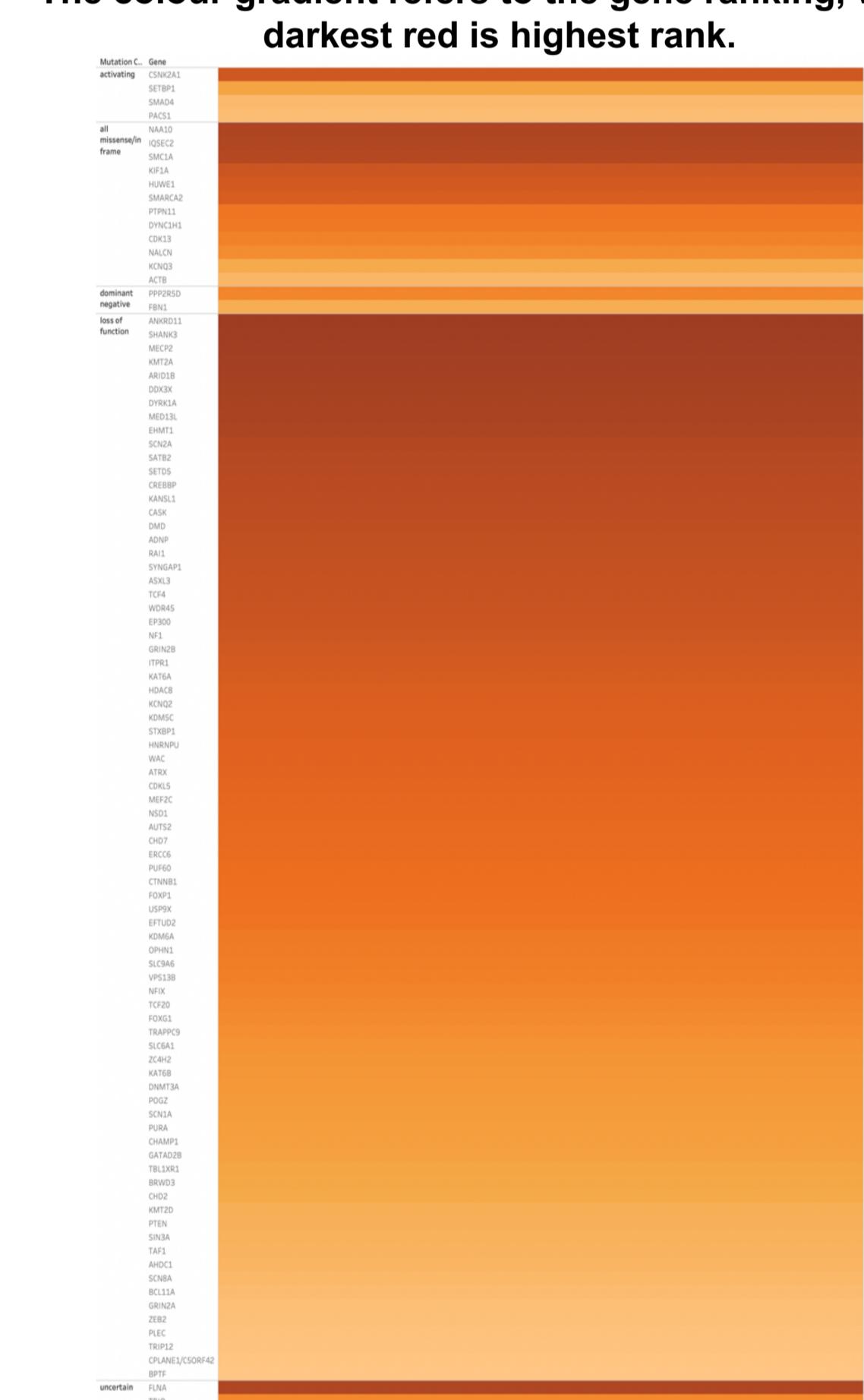
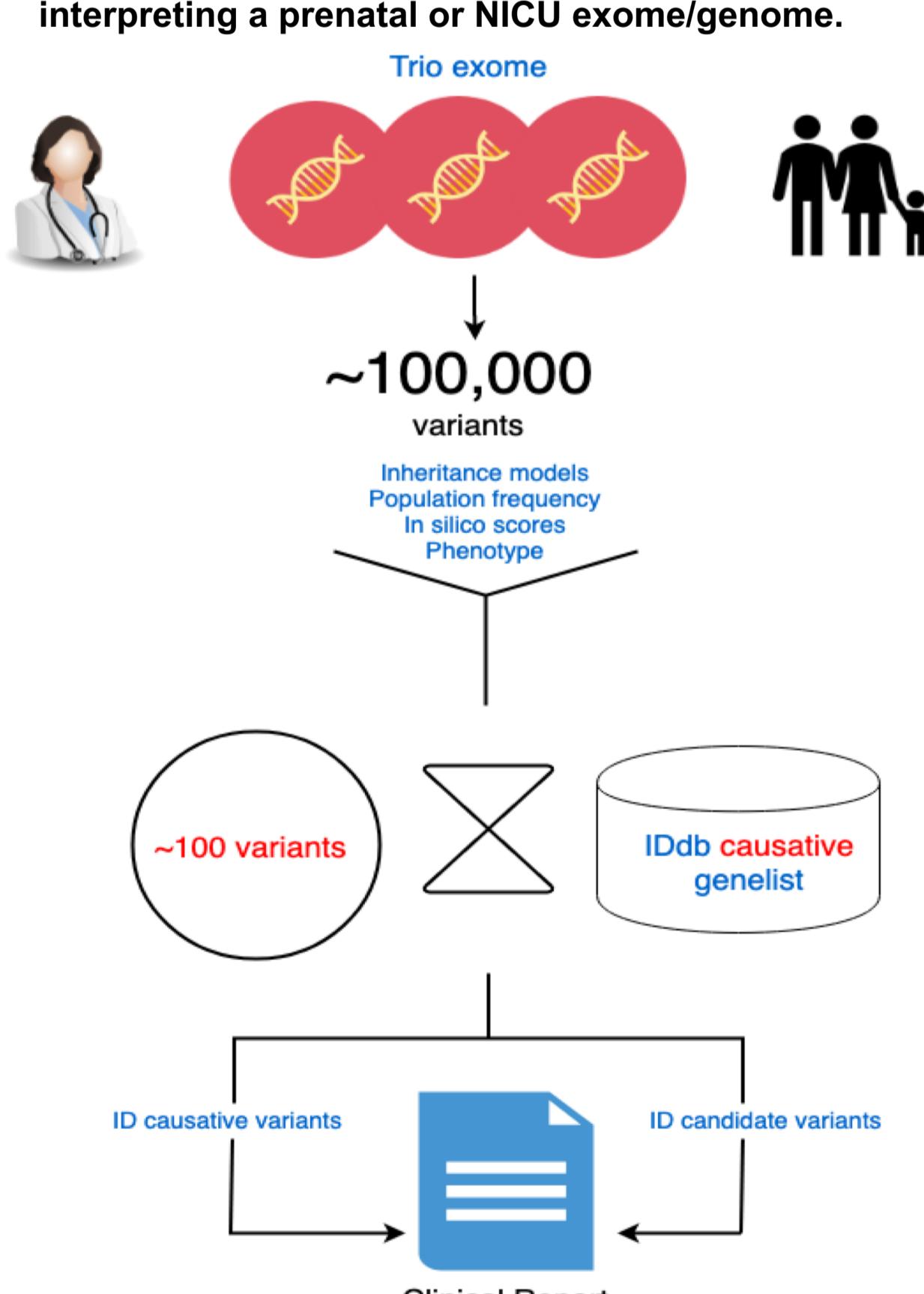


Figure 6: Mutation consequence of top 100 ID genes. The colour gradient refers to the gene ranking, where darkest red is highest rank.



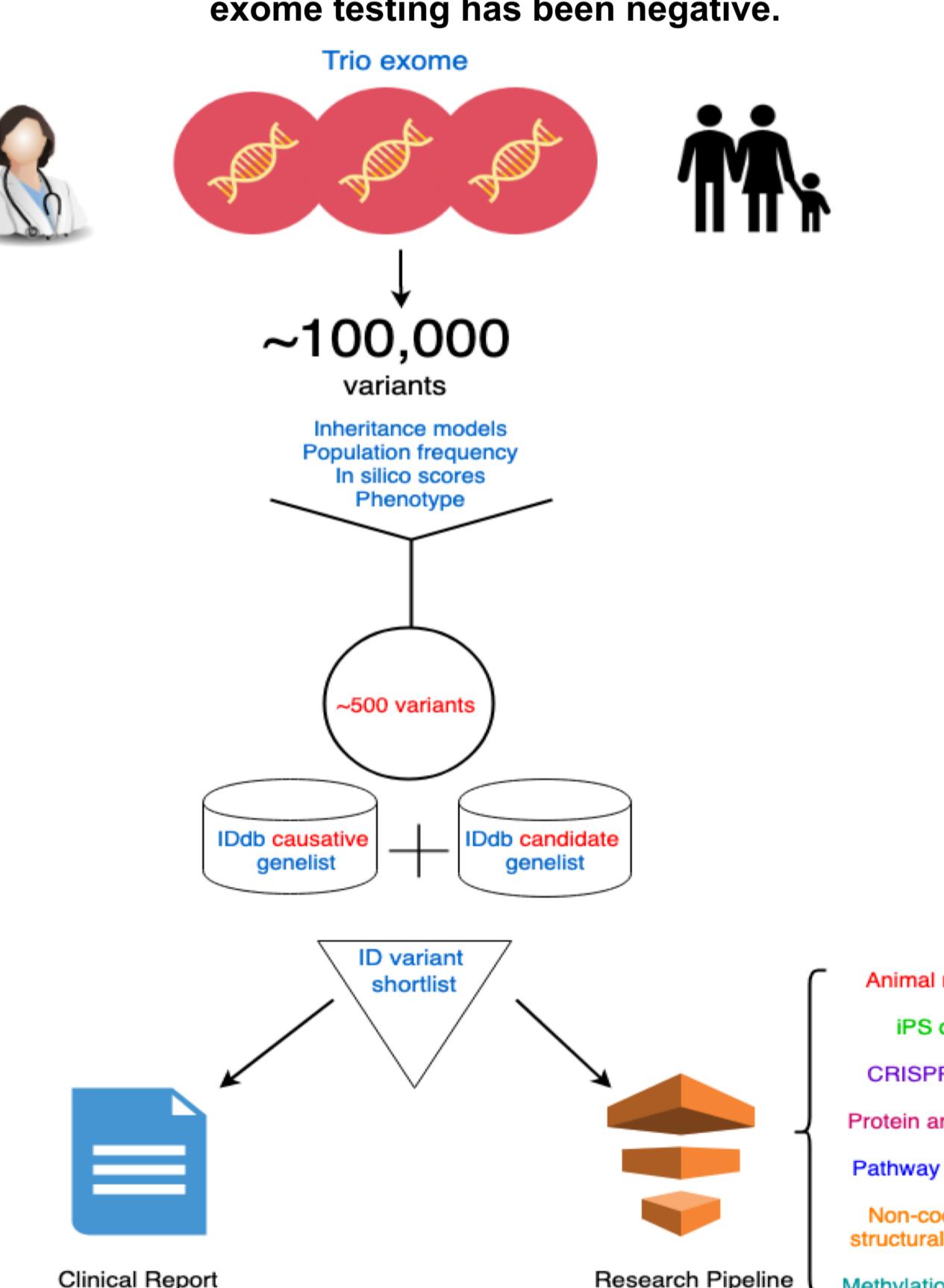
Clinical Application

Figure 7: IDdb can be used to rapidly prioritise genes to focus on when interpreting a prenatal or NICU exome/genome.



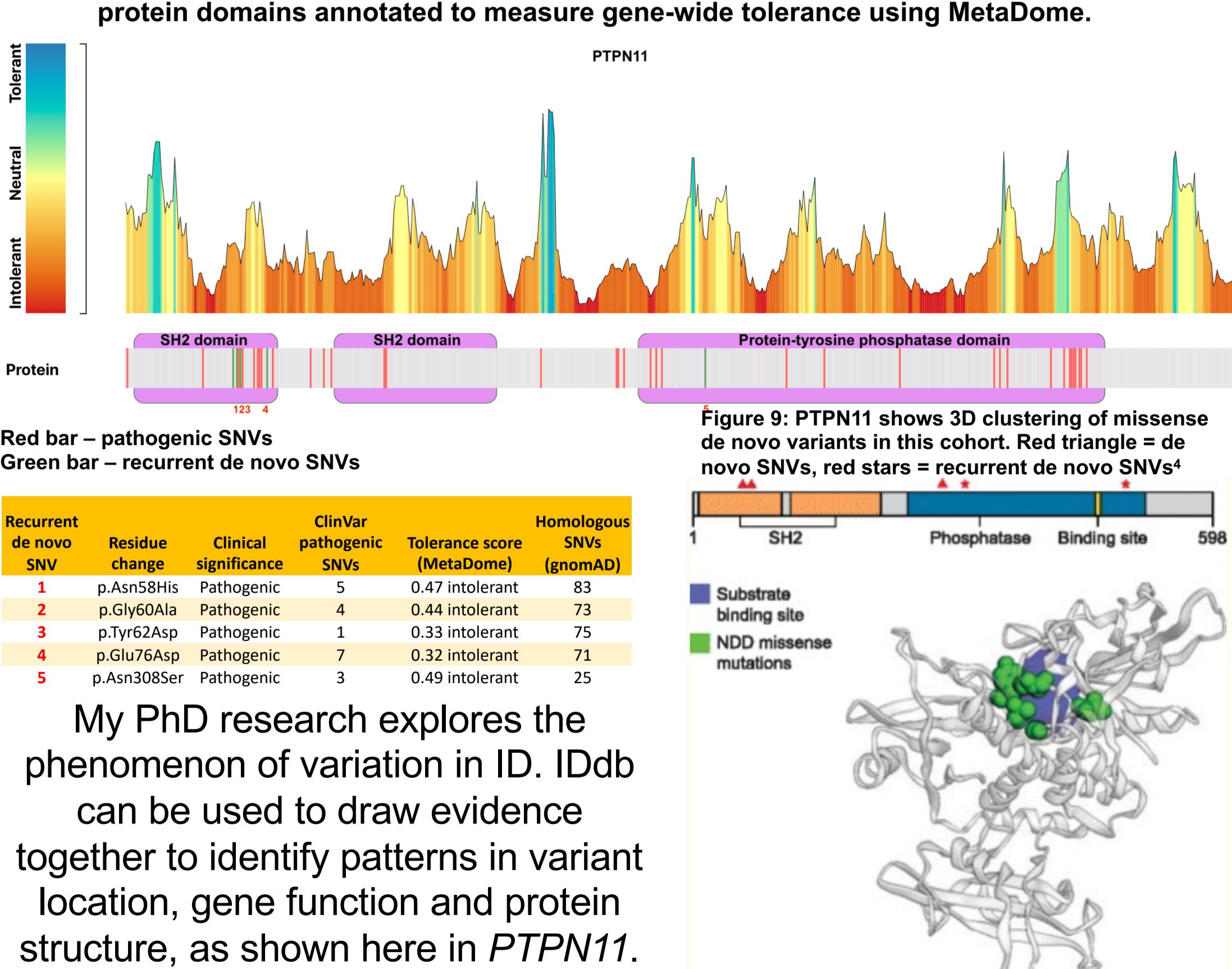
Translational Application

Figure 7: IDdb can be used to broaden the search within ID genes when prior exome testing has been negative.



Research Application

Figure 8: A schematic representation of the PTPN11 transcript of interest with Pfam protein domains annotated to measure gene-wide tolerance using MetaDome.



Conclusion

We present a concept for IDdb as a resource for intellectual disability genes from the literature. Standardising annotation from the literature will enable rapid variant curation. Integrating data from multiple databases will enable deeper insights into ID biology.

Future Directions

Phase 1: IDdb will be hosted on the website for the Centre for Research Excellence in Neurocognitive Disorders and updated at regular intervals.

Phase 2: ID gene variants will be curated, annotated and collated in the database.

Phase 3: Candidate ID genes and their variants will be added to the database.

Phase 4: IDdb data will be integrated using deep learning neural networks to predict ID subtype-gene associations.

References

- Disability expectations: Investing in a better life, a stronger Australia. PricewaterhouseCoopers, Nov 2011.
- How much does intellectual disability really cost? First estimates for Australia. Doran C et al. (2012). Journal of intell. & developmental disability, 37, 42-9.
- Gene Clinical Validity Curation Process SOP v6, August 2018. The Clinical Genome Resource. Gene Curation Working Group. <https://clinicalgenome.org/>
- Recurrent de novo mutations in neurodevelopmental disorders: properties and clinical implications. Wilfert AB et al. (2017) Genome medicine, 9(1), 101.