

CAGI slide guidelines

I agree to my slides being used according to the following terms:

Jochen Weile

- To disseminate the CAGI results, I encourage all CAGI participants to remix these CAGI slides for inclusion in their oral and poster presentations
- CAGI participants may use or publish any content of these slides only in compliance with the CAGI Data Use Agreement
- CAGI participants may not cite or publish any content of these slides except with the written permission of the originating author or in the primary CAGI publications
- CAGI participants must acknowledge the original authors of these slides and CAGI 2018. Include the acknowledgement banner across the bottom
- CAGI participants must include the CAGI credits slide in their presentations
- Slides with a red slash (no remix) may not be included in a presentation unless all attendees are registered CAGI participants who signed the data use agreement
- Slides with a red X over the entire slide may not be used in any circumstances
- Slides with a blue slash (embargoed) may not be used unless the embargo is explicitly lifted
- Slides without slash or X may be tweeted unless flagged with the “No tweet” icon.
- CAGI participants using these slides must explain the “No tweet” icon, which means that the slide content should not be disseminated outside the presenting conference hall. If the presentation venue permits pervasive tweeting, you may not include these slides in your talk



A growing atlas of variant effects

Jochen Weile, Roth Lab

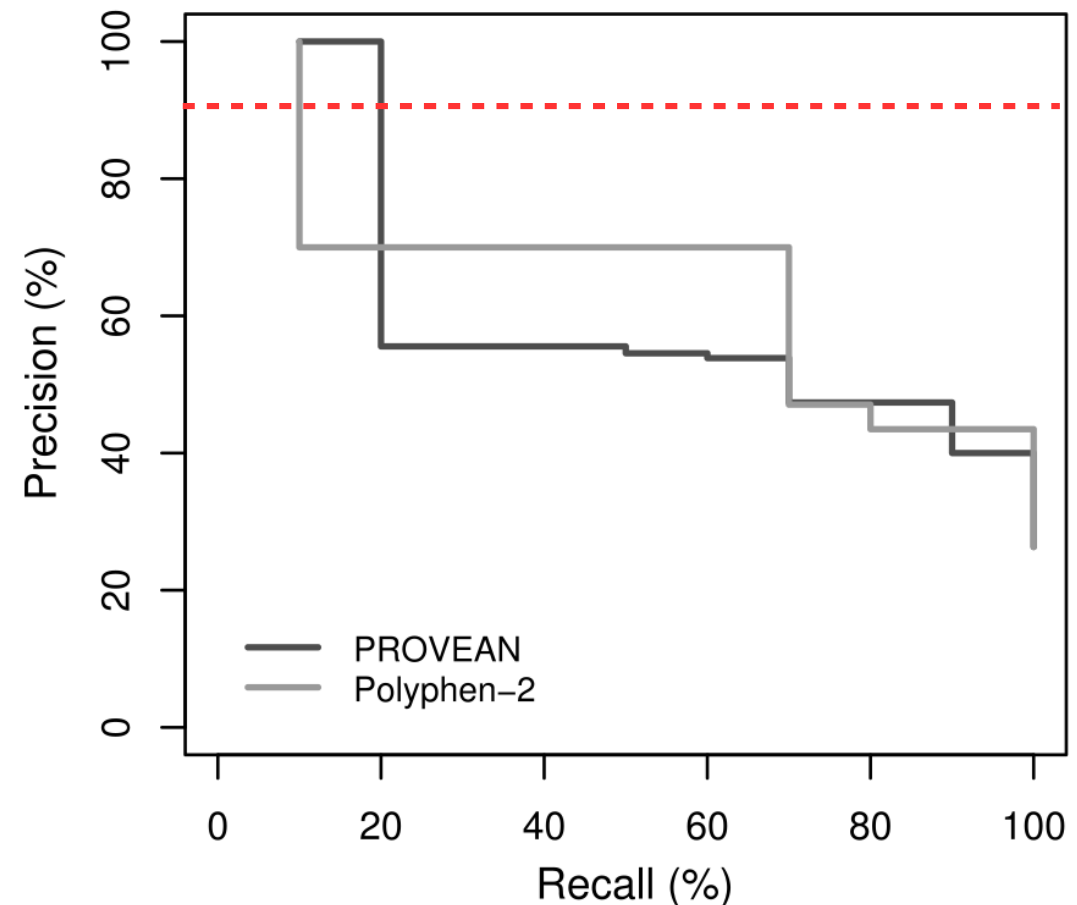
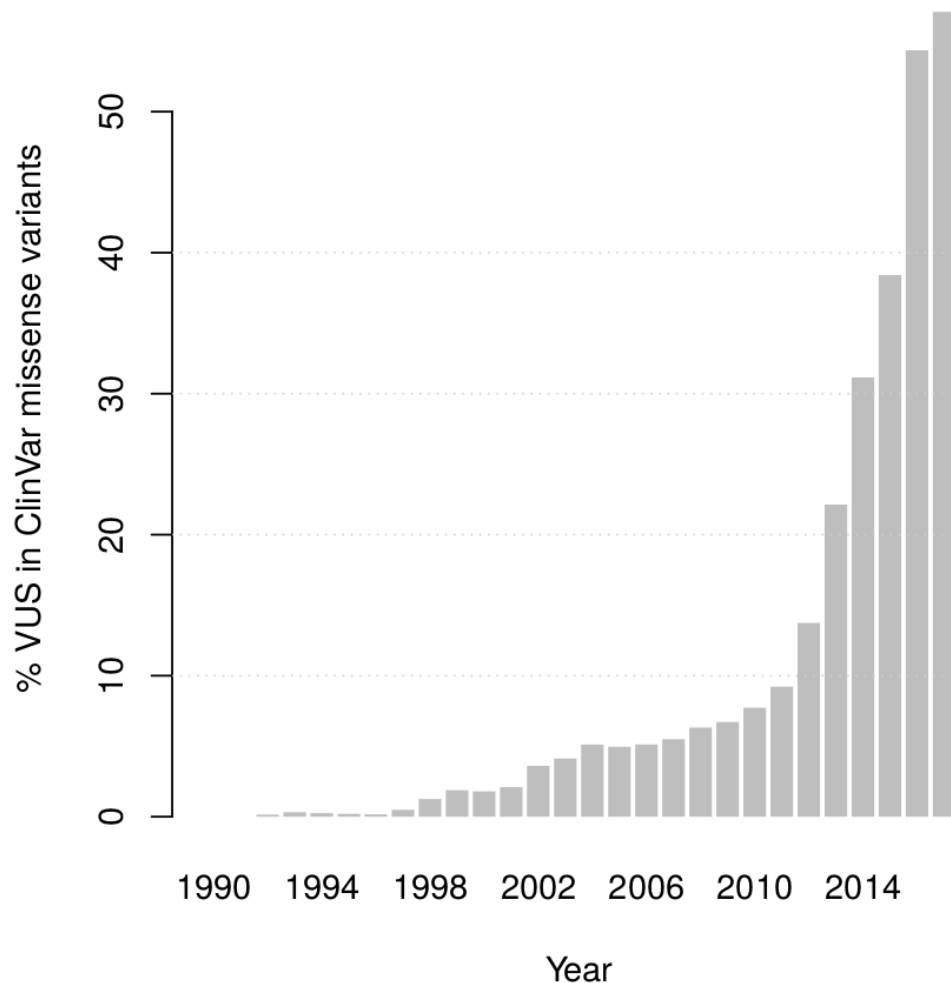
The Donnelly Centre, University of Toronto
Lunenfeld-Tanenbaum Research Institute, Mt Sinai Hospital, Toronto

CAGI 5 Conference 2018

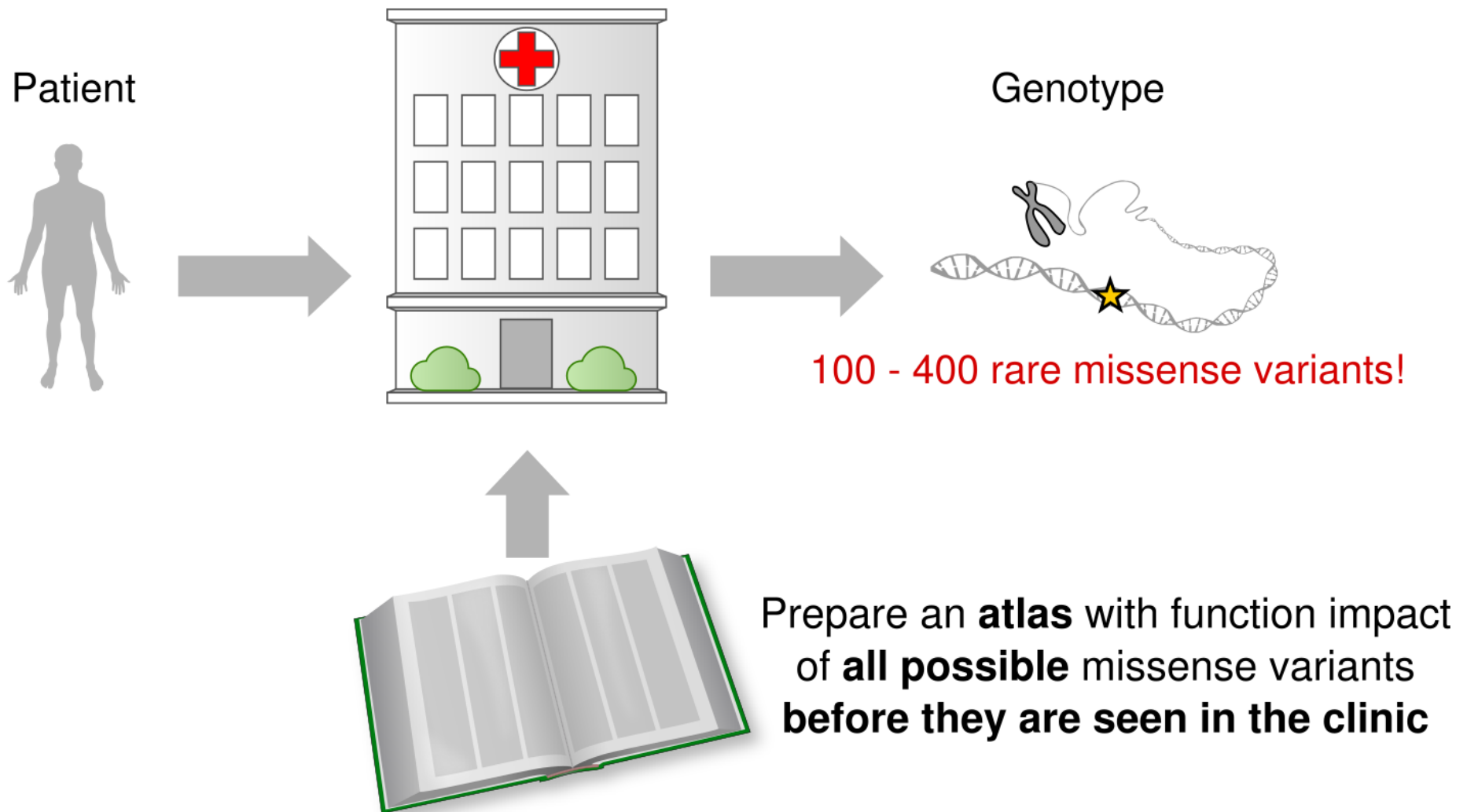
Overview

- **Problem:** Variants of Uncertain Significance
- **Approach:** Multiplex Assays of Variant Effect (MAVE)
- **Application 1:** Calmodulin and Heart disease
- **Application 2:** CBS, MTHFR and Homocysteinuria
- **Scaling up:** The MaveDB repository and ecosystem

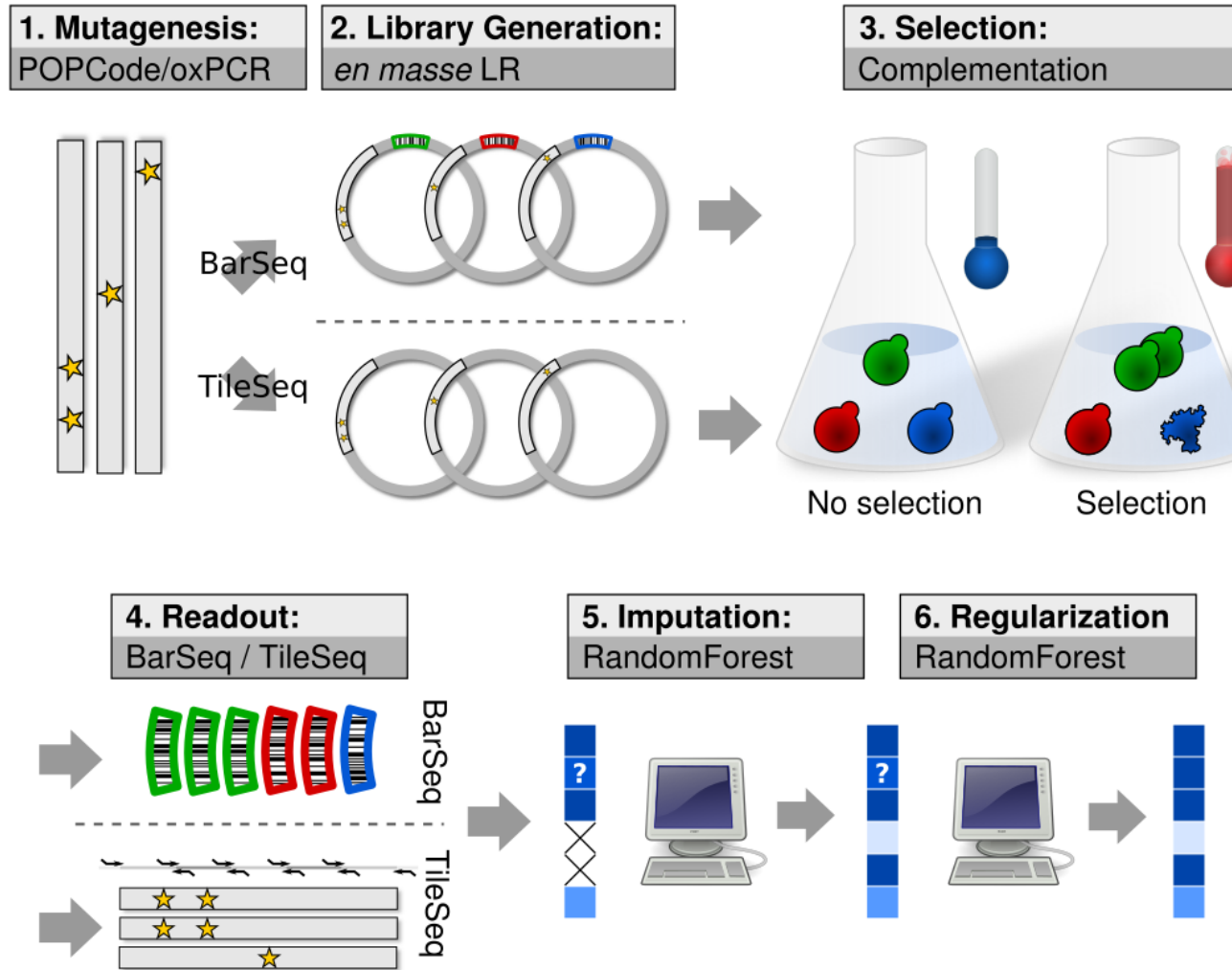
Variants of Uncertain Significance (VUS) are taking over



Proactive assaying of variants

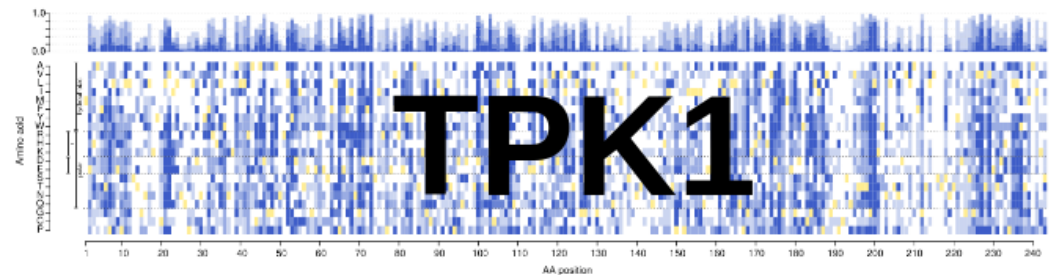
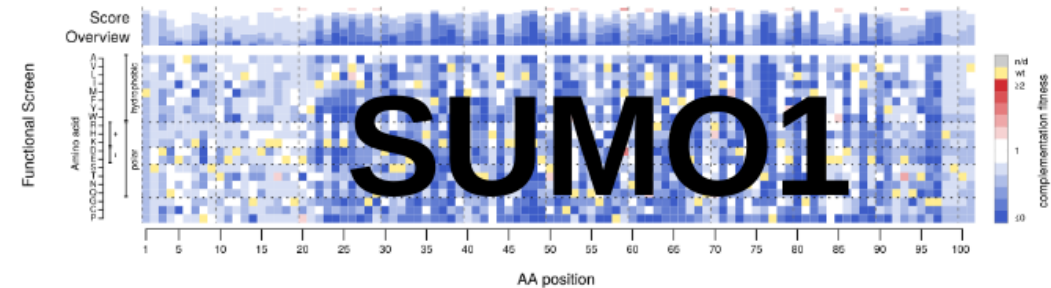
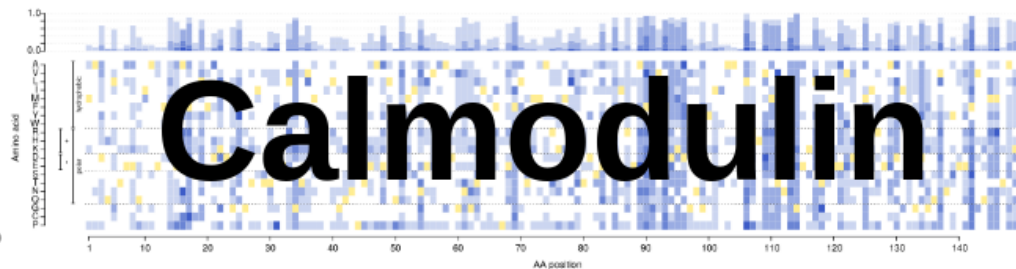
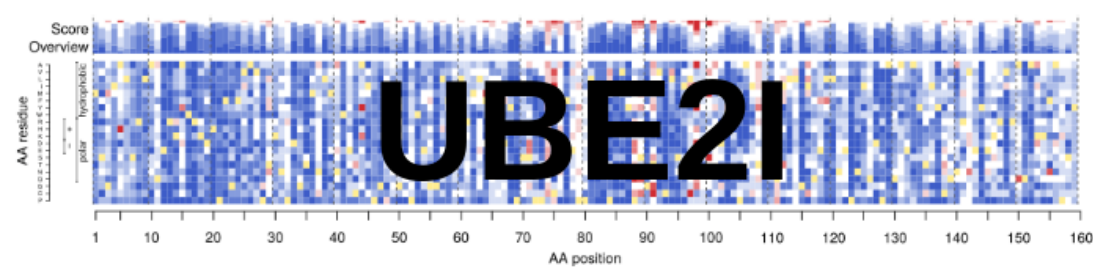


Multiplex Assays of Variant Effects (MAVE)



Pioneered by Doug Fowler, Stan Fields, Dan Bolon and others

Variant effect maps for four proteins



Now available in MSB



Published online: December 21, 2017

Method



molecular
systems
biology

A framework for exhaustively mapping functional missense variants

Jochen Weile^{1,2,3,4,†}, Song Sun^{1,2,3,4,5,†}, Atina G Cote^{1,2,3}, Jennifer Knapp^{1,2,3}, Marta Verby^{1,2,3}, Joseph C Mellor^{2,6}, Yingzhou Wu^{1,2,3,4}, Carles Pons⁷, Cassandra Wong^{1,2}, Natascha van Lieshout¹, Fan Yang^{1,2,3,4}, Murat Tasan^{1,2,3,4}, Guihong Tan^{2,3}, Shan Yang⁸, Douglas M Fowler⁹, Robert Nussbaum⁸, Jesse D Bloom¹⁰, Marc Vidal^{11,12}, David E Hill¹¹, Patrick Aloy^{7,13} & Frederick P Roth^{1,2,3,4,14,*}

Abstract

Although we now routinely sequence human genomes, we can confidently identify only a fraction of the sequence variants that have a functional impact. Here, we developed a deep mutational scanning framework that produces exhaustive maps for human missense variants by combining random codon mutagenesis and multiplexed functional variation assays with computational imputation and refinement. We applied this framework to four proteins corresponding to six human genes: UBE2I (encoding SUMO E2 conjugase), SUMO1 (small ubiquitin-like modifier), TPK1 (thiamin pyrophosphokinase), and CALM1/2/3 (three genes encoding the protein calmodulin). The resulting maps recapitulate known protein features and confidently identify pathogenic variation. Assays potentially amenable to deep mutational scanning are already available for 57% of human disease genes, suggesting that DMS could ultimately map functional variation for all human disease genes.

Keywords complementation; deep mutational scanning; genotype-phenotype; variants of uncertain significance

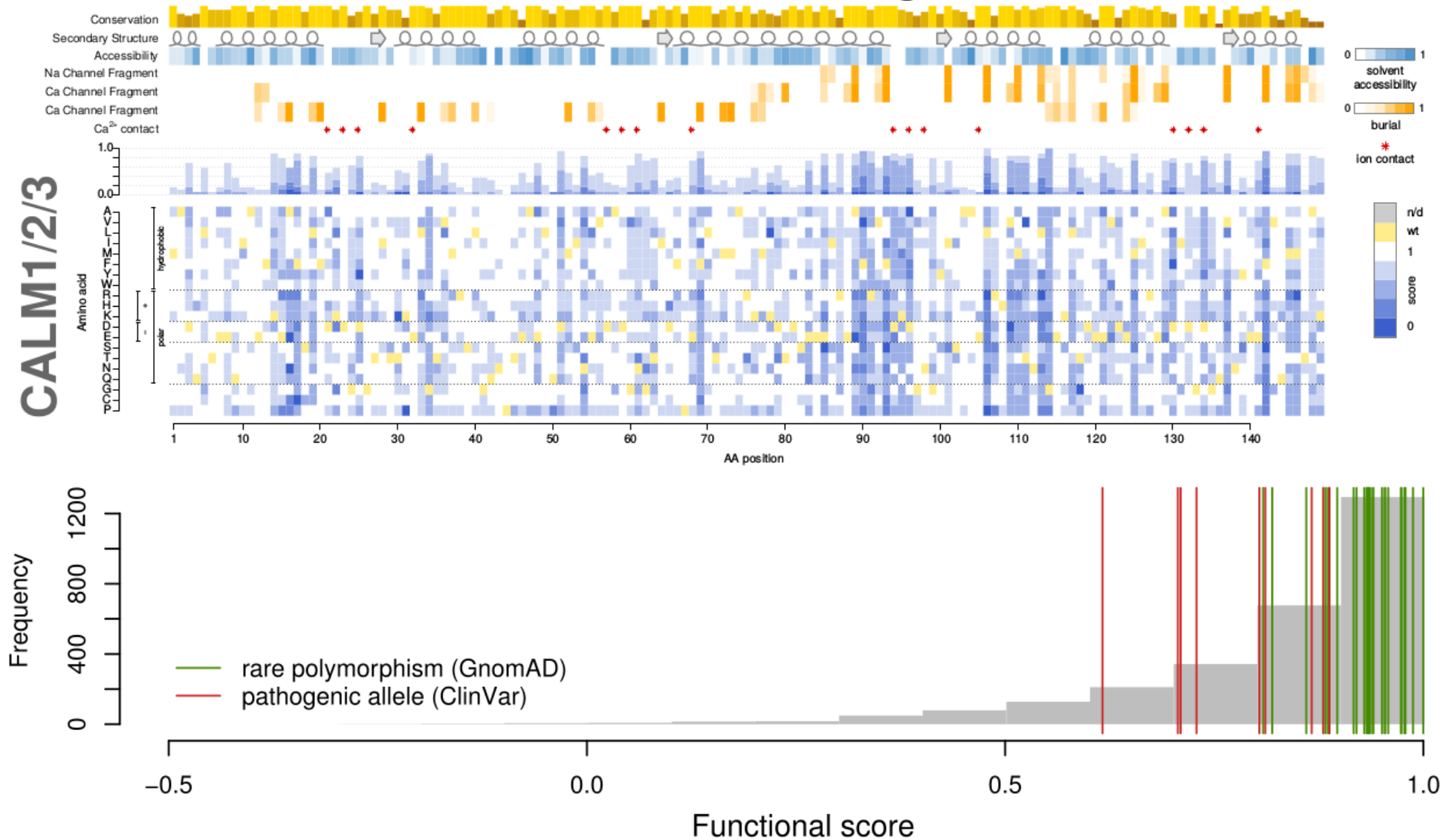
Subject Categories Chromatin, Epigenetics, Genomics & Functional Genomics; Genome-Scale & Integrative Biology; Methods & Resources
DOI 10.15252/msb.20177908 | Received 9 August 2017 | Revised 15 November 2017 | Accepted 18 November 2017
Mol Syst Biol. (2017) 13: 957

Introduction

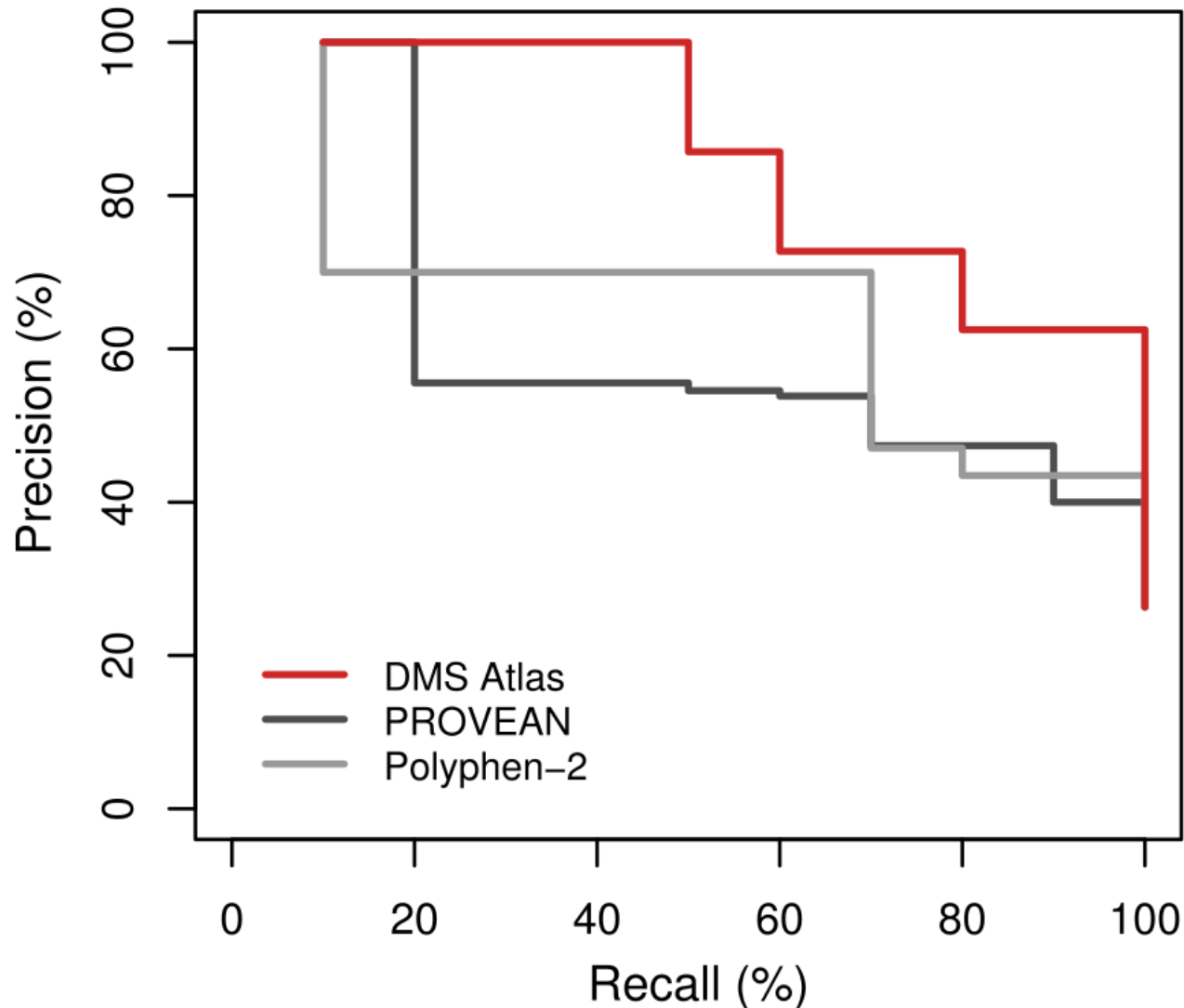
Millions of people will soon have their genomes sequenced. Unfortunately, we have only a limited ability to interpret personal genomes, each carrying 100–400 rare missense variants (The 1000 Genomes Project Consortium, 2015) of which many must currently be classified as “variants of uncertain significance” (VUS). For example, gene panel sequencing aimed at identifying germline cancer risk variants in families yielded VUS for the majority of missense variants (Maxwell *et al.*, 2016). Functional variants can be predicted, but when high precision is required, computational tools (Adzhubei *et al.*, 2010; Choi *et al.*, 2012) detect only one-third as many pathogenic variants as experimental assays (Sun *et al.*, 2016). Unfortunately, validated experimental assays enabling rapid clinical interpretation of variants are not available for the vast majority of human disease genes.

Deep mutational scanning (DMS) (Fowler *et al.*, 2010; Fowler & Fields, 2014; Starita *et al.*, 2017), a strategy for large-scale functional testing of variants, can functionally annotate a large fraction of amino acid substitutions for a substantial subset of residue positions. Recent DMS studies, for example, covered the critical RING domain of BRCA1 (Starita *et al.*, 2015) associated with breast cancer risk, and the PPARG protein associated with Mendelian lipodystrophy and increased risk of type 2 diabetes (Majithia *et al.*, 2016). Such maps can accurately identify functionality of a clinical variant in advance of that variant’s first clinical presentation. Diverse assays can be used for DMS (see Table EV1). Functional complementation

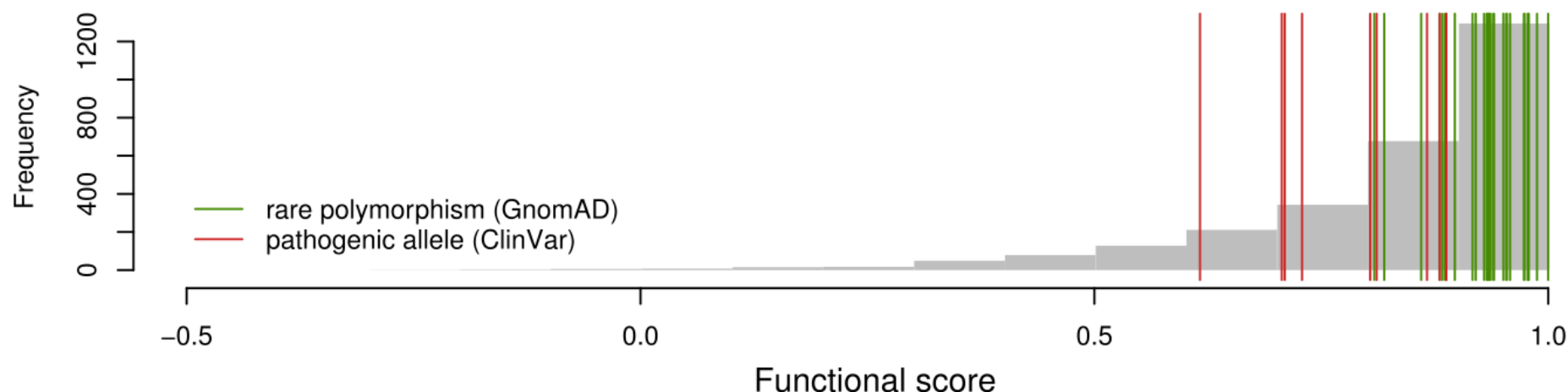
Calmodulin map distinguishes disease from benign variants



DMS map outperforms PROVEAN & PolyPhen-2

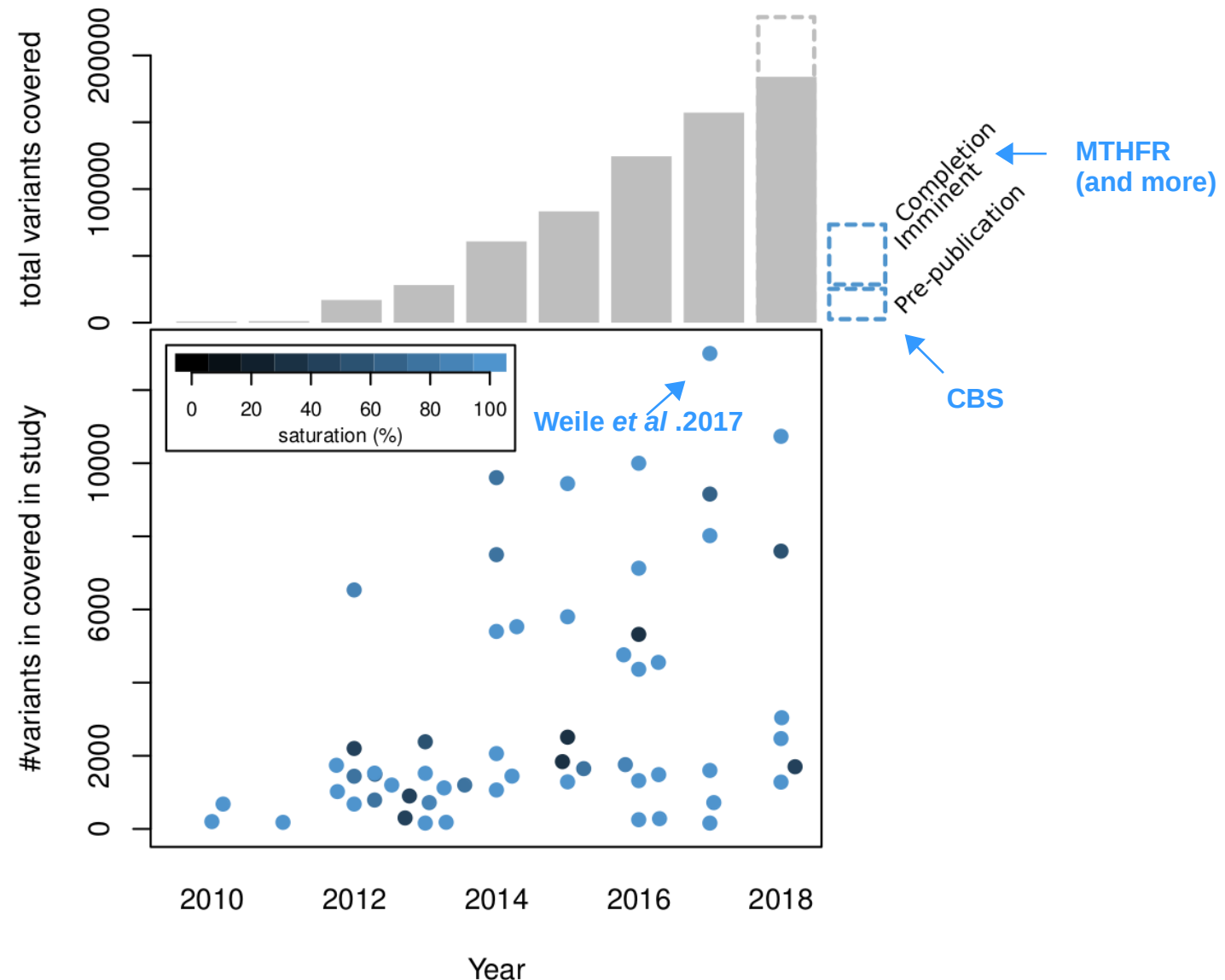


Invitae VUS predictions



Variant	MAF	sd/ rmsd	imp/ reg	pre-reg	DMS	DMS call	indication
D94A	NA	0.26	imputed	NA	0.46	likely damaging	Cardio
D96H	NA	0.26	imputed	NA	0.72	likely damaging	Cardio
I28V	$1 \cdot 10^{-5}$	0.05	mild	0.88	0.88	uncertain	Cardio
N98S	NA	0.05	mild	0.89	0.89	uncertain	Cardio
T35I	$4 \cdot 10^{-6}$	0.04	mild	0.93	0.93	likely benign	Non-Cardio
E48G	NA	0.05	mild	0.93	0.93	likely benign	Cardio
G26D	NA	0.06	mild	0.94	0.94	likely benign	Non-Cardio
T27S	$3 \cdot 10^{-5}$	0.05	mild	0.96	0.96	likely benign	Non-Cardio
V122A	NA	0.05	mild	0.98	0.98	likely benign	Non-Cardio
A104G	NA	0.08	mild	1.00	1.00	likely benign	Non-Cardio

Roth Lab to be largest contributor of variant effect data in 2018



MaveDB: An online repository for Variant Effect Maps

 MaveDB Home Search Documentation Profile Upload ▾ Logout 

Show entries Search:

Urn	Short description	Target name	Target organisms	Contributors
urn:mavedb:0000001-a-1	A joint Deep Mutational Scan of the human SUMO E2 conjugase UBE2I using functional complementation in yeast, combining DMS-BarSeq and DMS-TileSeq data, followed by machine-learning-based imputation and refinement.	UBE2I	<i>Homo sapiens</i>	Jochen Weile 
urn:mavedb:0000001-b-1	A machine-learning imputed and refined Deep Mutational Scan of the human SUMO	SUMO1	<i>Homo sapiens</i>	Jochen Weile 

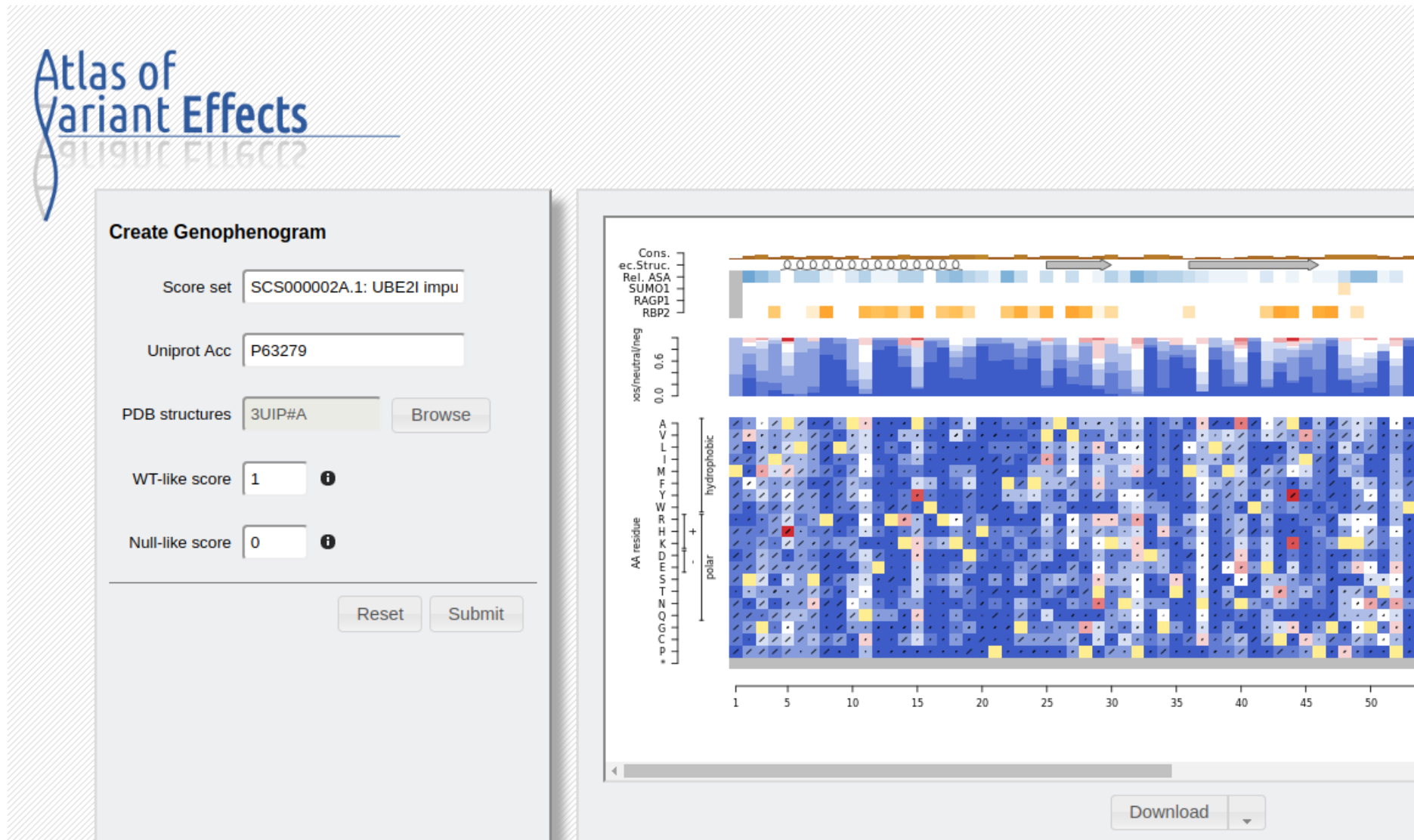


Alan Rubin



Doug Fowler

An ecosystem of services around MaveDB

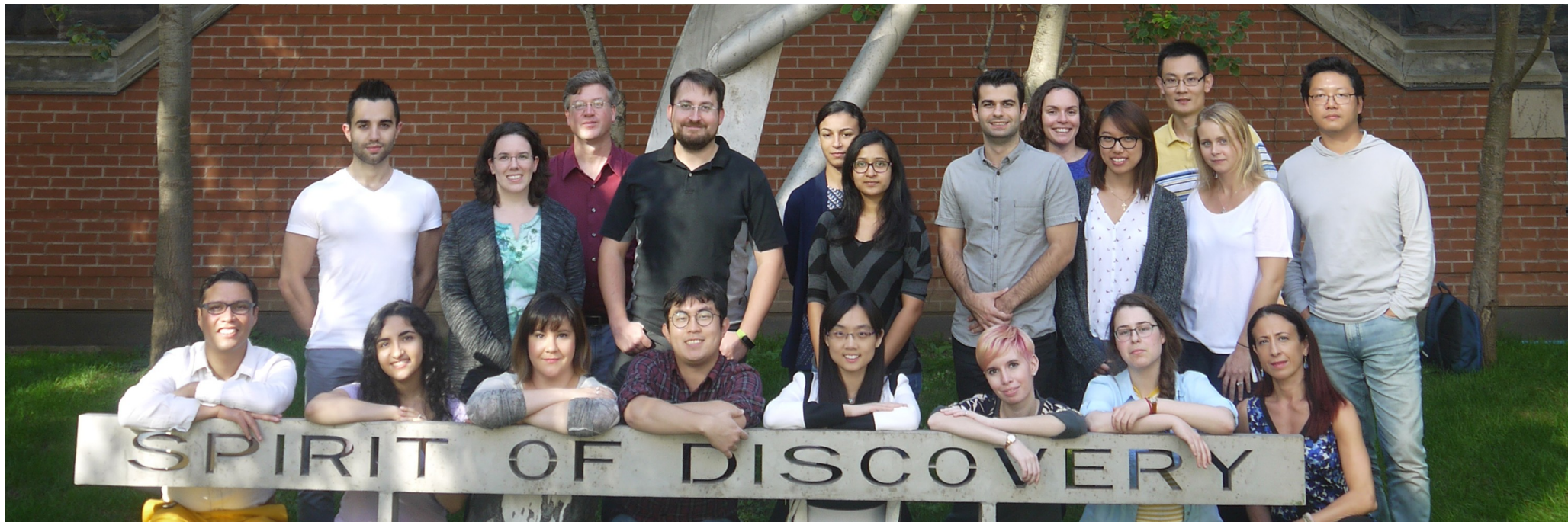


Summary

- Multiplexed Assays of Variant Effect (MAVEs) contribute to a growing genotype-phenotype atlas
- CBS and MTHFR MAVEs demonstrate clinical utility
- MaveDB will make MAVE data widely accessible

Thanks!

Fritz Roth, Jennifer Knapp, Hanane Ennajdaoui, Nishka Kishore, Song Sun, Marta Verby, Joe Wu, Atina Cote, Fan Yang, Natascha van Lieshout, Marinella Gebbia



Funding:

One Brave Idea
Canada Excellence Research Chair
NIH NHGCR CEGS program
Ontario Ministry of Research and Innovation

Collaborators:

Doug Fowler, Alan Rubin, Daniel Esposito, Jesse Bloom,
Guihong Tan, Charlie Boone, Brenda Andrews,
Shan Yang, Robert Nussbaum,
Victor Kozich