

CAGI slide pack guidelines

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

_____ *Panagiotis Katsonis* _____

Authors: please include THIS slide at the beginning of your presentation to indicate that you agree with reuse of your slides in CAGI presentations, according to the terms below, and digitally sign by writing your name above.

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 2013. 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

Note for slide authors: we expect all CAGI participants to abide by these restrictions, and will make best effort to ensure they are followed. However, but bear in mind that we have limited means of enforcing them, and therefore the restrictions cannot be guaranteed.



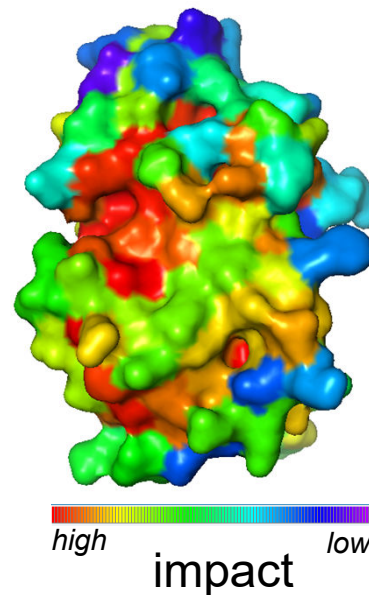
An Analytic Approach for the Impact of Protein Coding Variants

Panagiotis Katsonis, Rhonald Lua, and Olivier Lichtarge

Department of Molecular and Human Genetics



Evolutionary Action
(Katsonis and Lichtarge, 2014)



Protein:	SUMO ligase
Gene:	UBE2I
Length:	158 aa
Structure:	1A3S
Sequences:	137

Evolutionary Action Model

$$\Delta\varphi \approx \frac{\partial f}{\partial r_i} \cdot \Delta r_i$$

Phenotype response — $\Delta\varphi$

Evolutionary gradient — $\frac{\partial f}{\partial r_i}$

Genotype change — Δr_i

Evolutionary Action Model

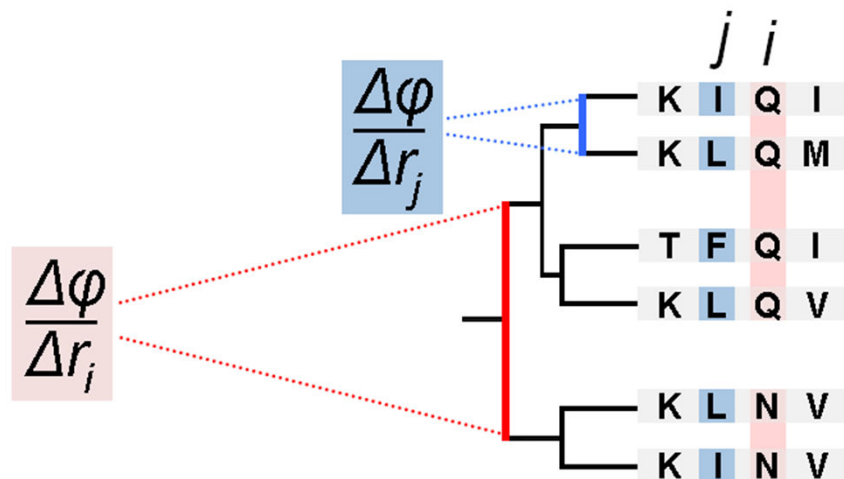
Phenotype response

$$\Delta\varphi \approx \frac{\partial f}{\partial r_i} \cdot \Delta r_i$$

Evolutionary gradient

Genotype change

Evolutionary trace importance



Evolutionary Action Model

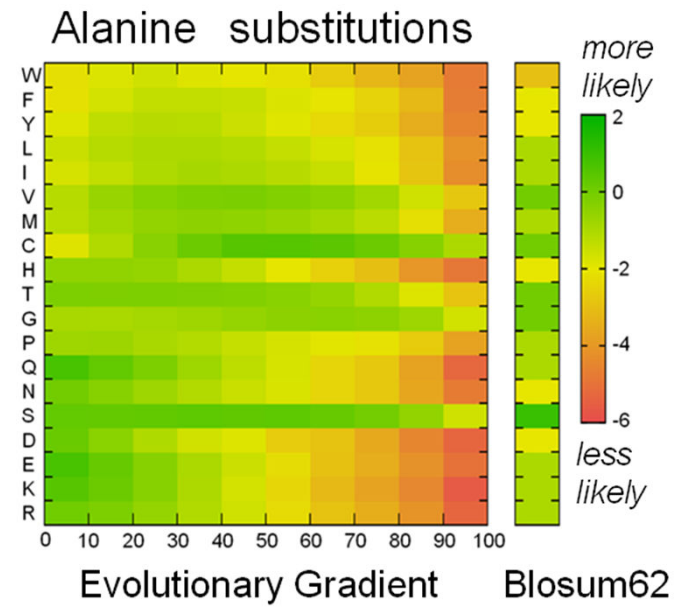
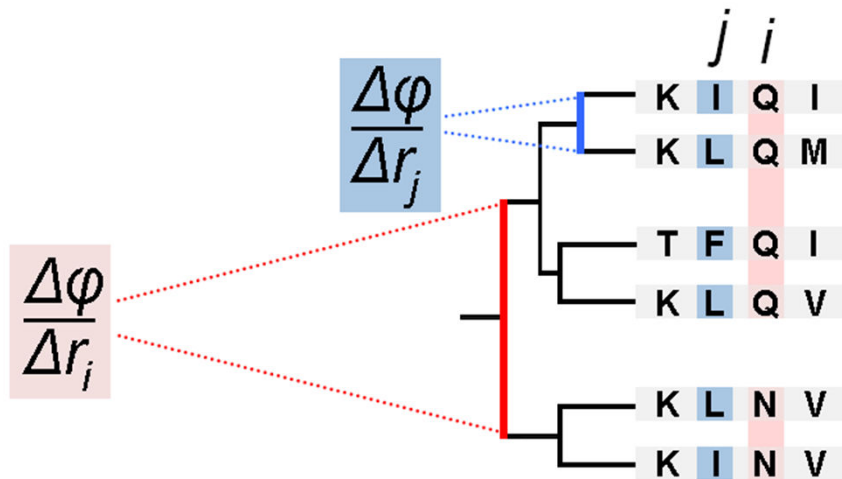
Phenotype response

$$\Delta\phi \approx \frac{\partial f}{\partial r_i} \cdot \Delta r_i$$

Evolutionary gradient

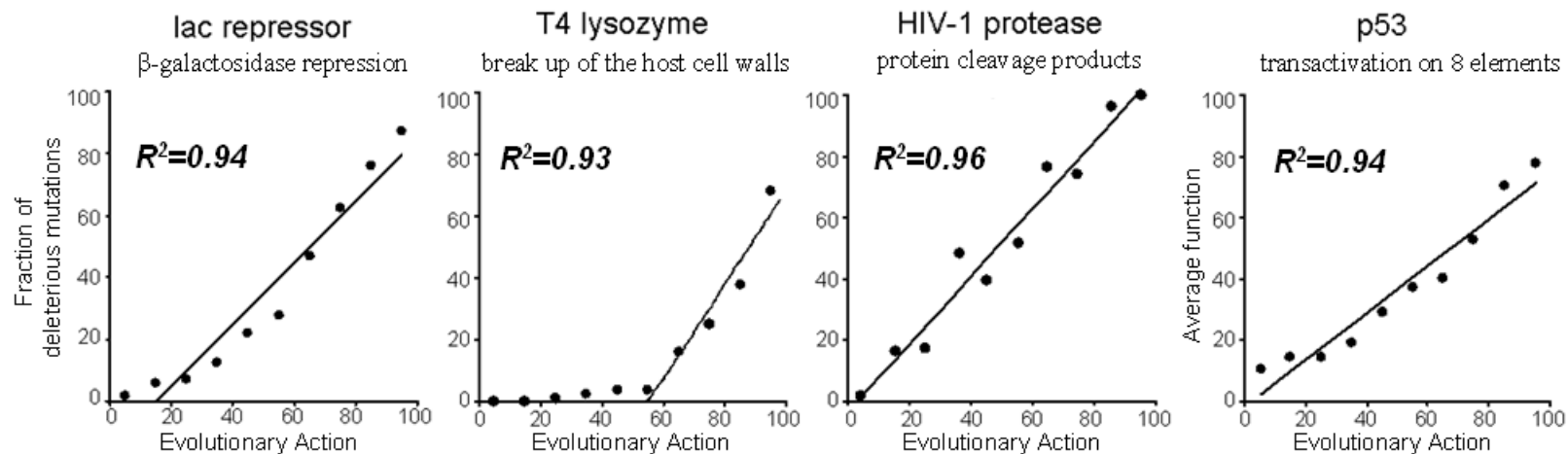
Genotype change

Evolutionary trace importance

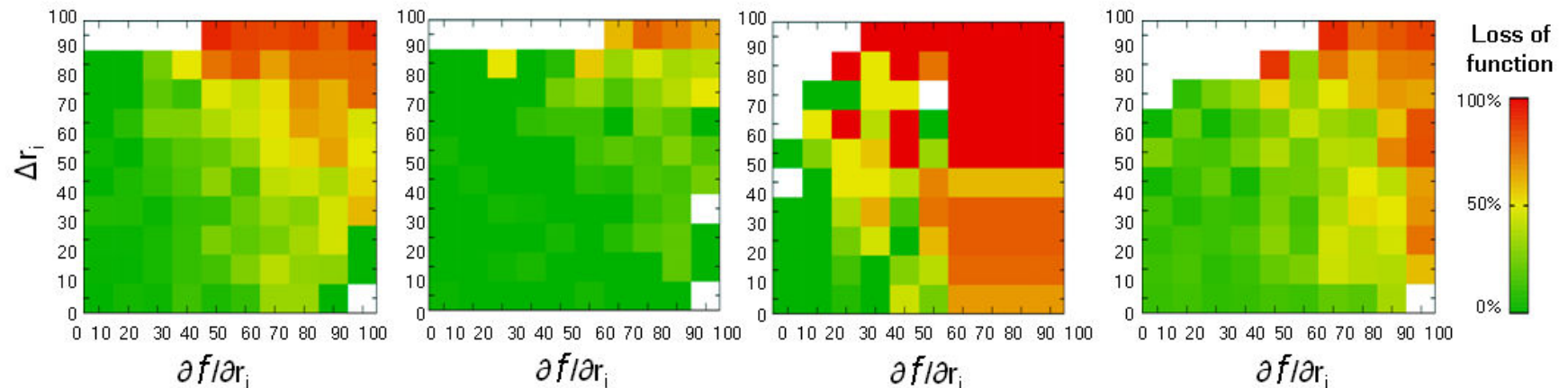


Validation of Evolutionary Action

The predicted phenotype change correlates to experimental assay data



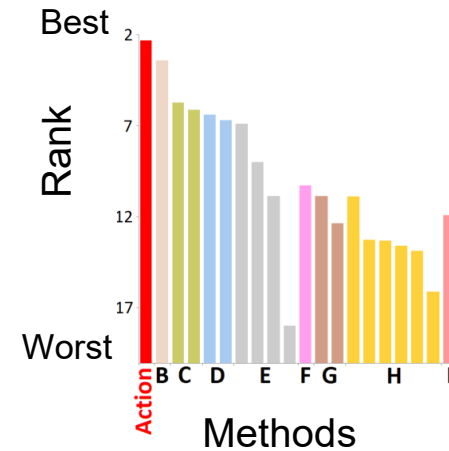
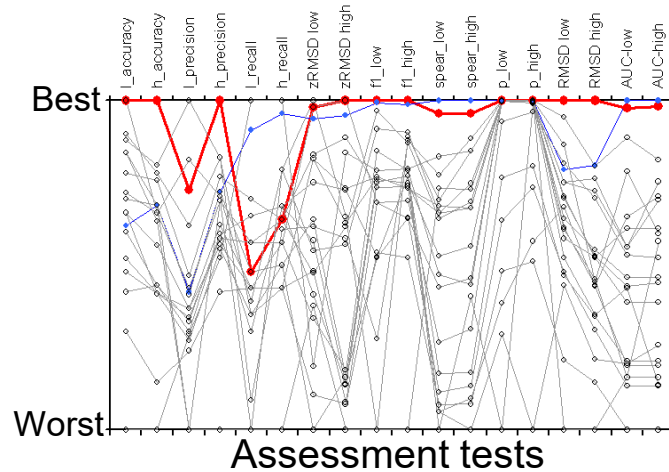
both terms of the equation contribute to the estimation of phenotype



E. Action at Past CAGI Contests

2011 Challenge:

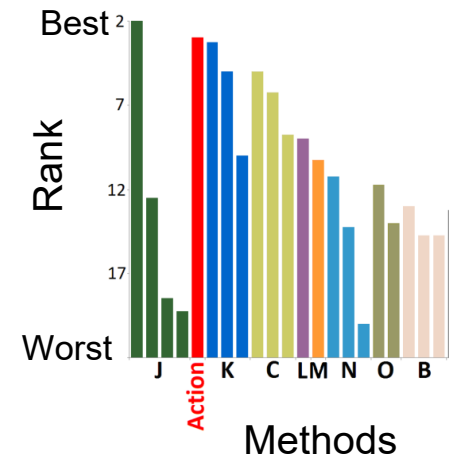
predict the function of 84 CBS mutants in high and in low pyridoxine concentration



2012-13 Challenge:

predict the cell proliferation for 10 variants of the p16 protein

Group name	Kendall	RMSD	AUC	#range 10	rank KCC	rank RMSD	rank auc	nk #range	avg rank	Overall rank
SPARKS-LAB	0.69	9.2	0.92	7	1	1	5	1	2	1
Lichtarge Lab	0.63	16.5	1	5	2	5	1	4	3	2
BioFoID	0.60	15.8	0.88	6	3	2	6	2	3.25	3
BioFoID	0.47	15.9	0.84	6	7	3	8	2	5	4
moult_lab	0.51	17.7	0.96	4	4	6	3	7	5	4
moult_lab	0.51	18.8	0.96	3	4	9	3	9	6.25	6
moult_lab	0.38	20.2	0.88	3	9	11	6	9	8.75	7
Anonymous1	0.45	23.5	1	3	8	18	1	9	9	8
BioFoID	0.28	16.4	0.72	2	11	4	10	15	10	9
FoxChase	0.20	18.2	0.68	4	14	7	13	7	10.25	10
Gough	0.07	20.3	0.76	3	15	12	9	9	11.25	11
LU bioinfo	0.05	20.1	0.42	5	16	10	17	4	11.75	12
SPARKS-LAB	0.20	18.5	0.4	3	13	8	20	9	12.5	13
ySNAP?	0.25	21.8	0.72	2	12	15	10	15	13	14
BolognaBiocomputing	0.47	25.5	0.7	2	6	20	12	15	13.25	15
LU bioinfo	0.34	25.7	0.58	3	10	21	16	9	14	16
Gough	-0.40	23.3	0.42	5	19	17	17	4	14.25	17
ySNAP?	0.02	21.3	0.62	2	17	13	14	15	14.75	18
ySNAP?	0.02	21.3	0.62	2	17	13	14	15	14.75	18
SPARKS-LAB	-0.56	22.5	0.08	2	21	16	22	15	18.5	20
SPARKS-LAB	-0.60	23.7	0.12	2	22	19	21	15	19.25	21
Gough	-0.40	39.2	0.42	1	19	22	17	22	20	22



SUMO Ligase Challenge Approach

Predict the competitive growth score (C) of:

Challenge 1:
high-accuracy subset of
215 single amino acid variants

$$C = 1 - EA/100$$

Challenge 2:
remaining
463 single amino acid variants

Accounting for GOF mutants:

when wild-type amino acid appears less frequently
than the substitute in the protein family alignment

$$C = 1 + EA/100$$

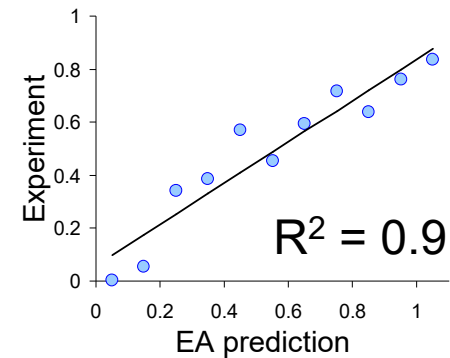
Challenge 3:
4,427 alleles corresponding to clones
with two or more amino acid variants

$$C_A = C_1 \cdot C_2 \cdot \dots \cdot C_N$$

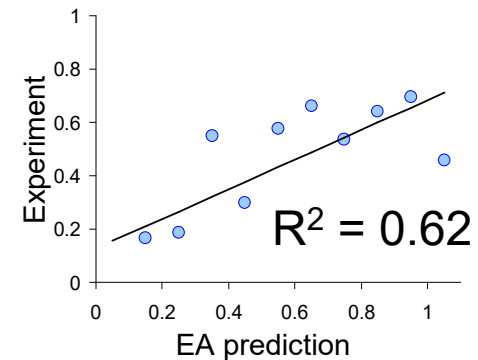
SUMO Ligase Challenge Results

Predict the competitive growth score (C) of:

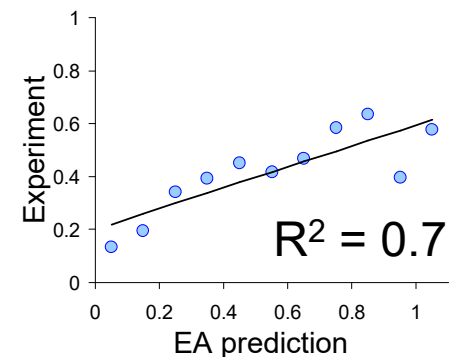
Challenge 1:
high-accuracy subset of
215 single amino acid variants



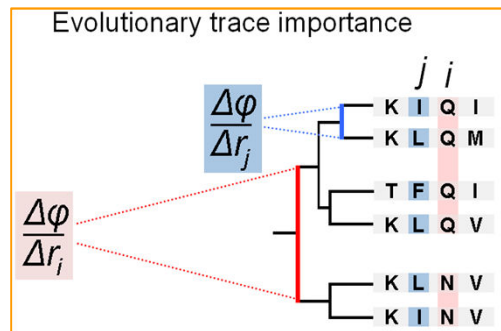
Challenge 2:
remaining
463 single amino acid variants



Challenge 3:
4,427 alleles corresponding to clones
with two or more amino acid variants



Summary



Phenotype
response

$$\Delta\phi \approx \frac{\partial f}{\partial r_i} \cdot \Delta r_i$$

