

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 6. 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

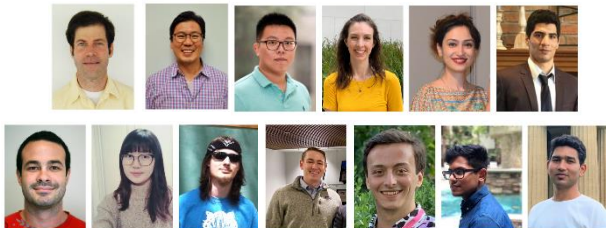


Rare Genomes Project (RGP): Evolutionary Action Suggests Causal Variants



Panagiotis Katsonis, Amanda Williams, and Olivier Lichtarge

Department of Molecular and Human Genetics, Baylor College of Medicine



Lichtarge lab



Baylor College of Medicine



Texas Medical Center, Houston TX

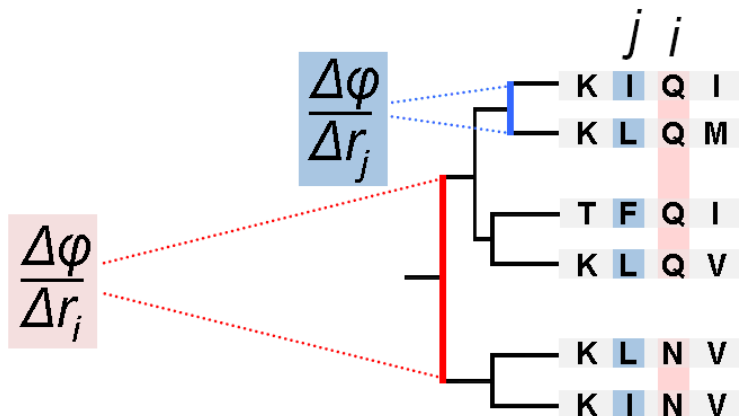
Evolutionary Action Model

Phenotype $\varphi = f(\gamma)$ Genotype

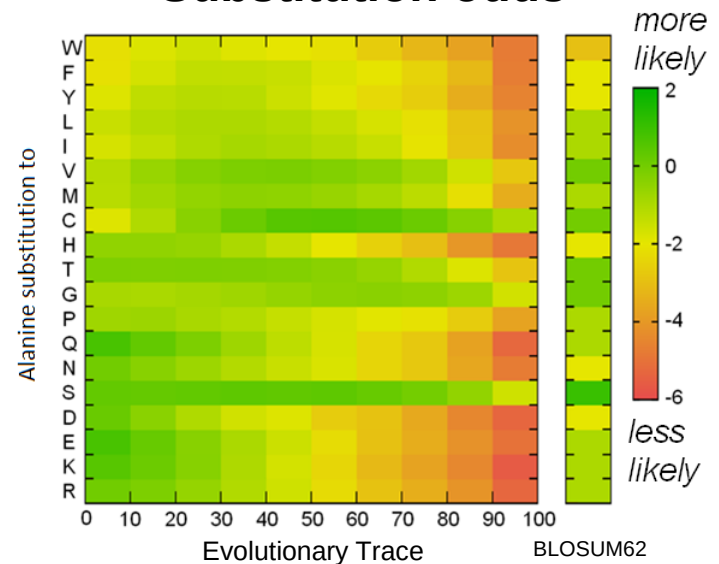
Phenotype change $\Delta\varphi \approx \frac{\partial f}{\partial \gamma_i} \cdot \Delta\gamma_i$ Genotype change

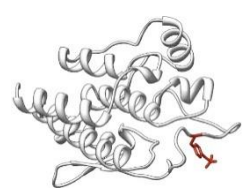
Evolutionary gradient

Evolutionary Trace Ranks



Substitution odds





CAGI Challenges

CAGI 1

- CBS

CAGI 2

- CBS

CAGI 3

- p16
- MRN
- BRCA

CAGI 4

- SUMO ligase
- Pyruvate kinase
- NAGLU

CAGI 5

- CALM1
- TPMT and PTEN
- GAA
- Enigma
- Clotting_disease
- SickKids5

CAGI6 (2021-2)

variants

- MTHFR
- HMBS
- STK11

clinic

- Sherlock

folding

- MAPK1
- MAPK3
- CaM

exomes

- Intellectual Disability
- SickKids
- Rare Genomes Project

Prediction Challenge

Identify one or more causal variants for each proband of 30 families

The clinical phenotype for each proband was given in the form of Human Phenotype Ontology (HPO) terms (multiple developmental or other traits per proband)

Test data:

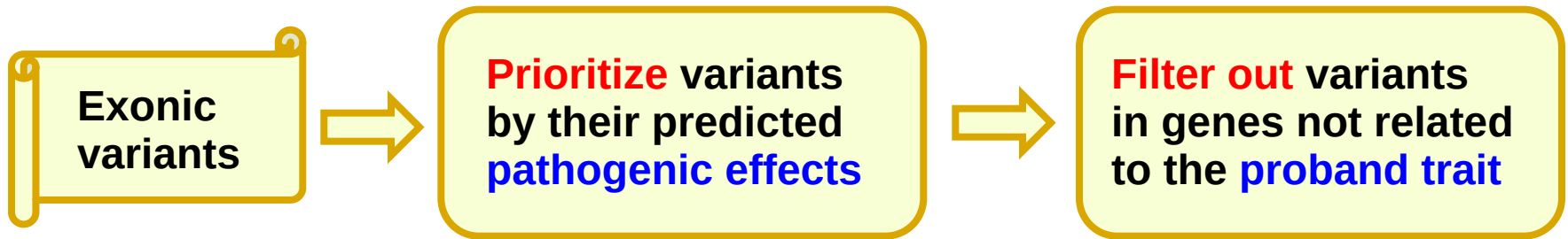
- 23 trio families
- 2 quad families (parents and two affected siblings)
- 2 probands with no parents
- 2 probands with father only
- 1 proband with mother only

Training data:

- 26 trio families
- 0 quad families (parents and two affected siblings)
- 6 probands with no parents
- 0 probands with father only
- 3 proband with mother only

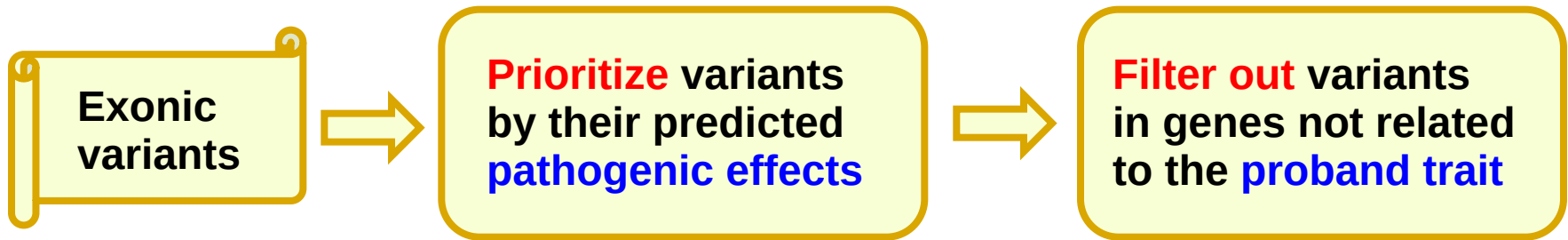
Approaches

MODEL 1

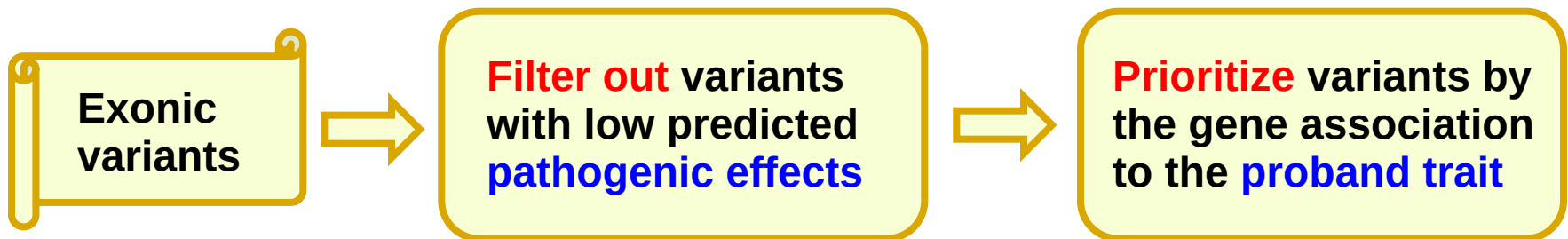


Approaches

MODEL 1



MODELS 2 & 3



HYPOTHESIS 1:

THERE IS ONE DE NOVO CAUSAL VARIANT

De novo variant: *found in probands, not found in their parents*

Pathogenic:

SCORE_{De novo}

- *large Evolutionary Action score*
- *rare or unseen in population*
- *true call*
- *gene is intolerant to variants*

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



PASS, DP, AD

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



Trait associations:

TRAIT

- *HPO*
- *ClinVar*
- *DisGeNet*
- *HumSavar*



HYPOTHESIS 1: THERE IS ONE DE NOVO CAUSAL VARIANT

Examples :

RGF_TEST_4_PROBAND

Chrom	POS	REF	MUT	GENE	SUB	read_Proband	read_Parents	Filter	call_counts	call_ratio	-log(GnomAD_AF)	train_AF	test_AF	Gene_intol	EA	SCORE	TRAIT
13	101229500	C	A	NALCN	G478W	15/34;	0/34;0/23	PASS	15	0.441	6.000	0.000	0.006	90.33	92.42	63.3009	2
22	31344951	C	T	PATZ1	A218T	15/41;	0/32;0/35	PASS	15	0.366	6.000	0.000	0.006	97	24.37	15.97025	0
9	104605393	G	A	OR13C2	P79S	3/27;	0/36;0/42	PASS	3	0.111	4.915	0.011	0.006	23.61	82.52	0.318219	0
16	14886779	G	A	NOMO1	E1081K	11/48;	0/37;0/33	PASS	11	0.229	2.224	0.033	0.029	47.42	30.87	0.208553	0
7	100996823	C	T	MUC12	T2087M	9/43;	0/12;0/52	PASS	9	0.209	2.679	0.039	0.029	62.65	54.38	0.100904	0
9	104605423	A	T	OR13C2	L69M	3/30;	0/36;0/42	PASS	3	0.100	3.763	0.006	0.006	23.61	20.44	0.072624	0
19	22087991	A	T	ZNF257	I5F	4/36;	0/41;1/37	PASS	4	0.111	6.000	0.056	0.041	38.77	42.16	0.047781	0
3	195725943	C	T	MUC20	T447M	4/30;	3/33;2/36	PASS	4	0.133	2.209	0.072	0.059	26.96	64.46	0.025498	0
7	101036923	A	G	MUC17	N1836S	5/40;	0/38;0/56	PASS	5	0.125	6.000	0.039	0.012	59.2	7.21	0.025354	0
9	104605372	G	C	OR13C2	L86V	2/27;	0/37;0/48	PASS	2	0.074	4.016	0.011	0.006	23.61	18.06	0.017331	0
5	181125151	G	T	OR2V1	L52M	7/43;	0/36;0/39	PASS	7	0.163	4.497	0.000	0.006	5.04	0.23	0.011743	0
19	22087997	G	A	ZNF257	E7K	4/36;	0/41;1/37	PASS	4	0.111	4.986	0.111	0.088	38.77	13.06	0.005968	0
14	104948916	A	G	AHNAK2	S2179P;	2/27;	0/16;0/41	PASS	2	0.074	3.104	0.006	0.012	38.69	13.08	0.004115	0
19	22087981	G	C	ZNF257	M45I	3/35;	0/43;0/41	PASS	3	0.086	5.256	0.050	0.041	38.77	4.87	0.002537	0
11	4946601	T	C	OR51A4	N167S	4/36;	0/23;0/34	PASS	4	0.111	1.107	0.289	0.176	12.91	11.17	0.000302	0

RGF_TEST_8_PROBAND

Chrom	POS	REF	MUT	GENE	SUB	read_Proband	read_Parents	Filter	call_counts	call_ratio	-log(GnomAD_AF)	train_AF	test_AF	Gene_intol	EA	SCORE	TRAIT
1	248638301	T	TCA	OR2T35	fs-indel	31/31;	0/29;0/11	PASS	31	1	2.804424114	0.75	0.8229	15.86	100	0.614064	0
19	8499639	T	G	PRAM1	K57Q	15/36;	0/44;0/29	PASS	15	0.41667	1.332536774	0.1067	0.1304	53.33	16.84	0.071475	0
7	101034124	G	A	MUC17	R903H	20/78;	0/57;0/27	PASS	20	0.25641	1.56185178	0.1111	0.0824	59.2	13.28	0.038063	0
7	101034097	C	A	MUC17	T894K	20/80;	0/60;0/27	PASS	20	0.25	1.573425307	0.1167	0.0824	59.2	9.14	0.024388	0
3	75741673	C	T	ZNF717	D41N	6/50;	2/35;0/34	PASS	6	0.12	2.096589296	0.0444	0.0647	-	18.99	0.010715	0
14	94379526	C	T	SERPINA	V335I	3/35;	0/37;0/38	PASS	3	0.08571	5.400434581	0.0389	0.0471	49.04	16.16	0.008281	0
6	29944148	C	T	HLA-A	H216Y	9/33;	0/42;0/2	PASS	9	0.27273	1.27395206	0.0611	0.0529	87.8	48.86	0.007307	0
19	4511373	T	C	PLIN4	N849D	3/35;	0/38;0/27	PASS	3	0.08571	3.172014621	0.0667	0.0765	42.17	31.44	0.004146	0
19	21294735	C	G	ZNF708	M77I	3/41;	0/50;0/31	PASS	3	0.07317	6	0.0222	0.0353	26.92	3.09	0.001605	0
7	144185604	A	G	CTAGE8	I701V	20/60;	0/39;0/46	PASS	20	0.33333	1.180300799	0.1207	0.1765	-	0.35	0.001205	0

HYPOTHESIS 2: THERE IS ONE INHERITED CAUSAL VARIANT

X-linked variant: male proband, heterozygous in mother, not in father

Pathogenic:

- *large Evolutionary Action score*
- *rare or unseen in population*
- *true call*
- *gene is intolerant to variants*

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



PASS, DP, AD

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



SCORE_{X-linked}

Trait associations:

- *HPO*
- *ClinVar*
- *DisGeNet*
- *HumSavar*



TRAIT

HYPOTHESIS 2: THERE IS ONE INHERITED CAUSAL VARIANT

Examples :

RGP_TEST_9_PROBAND

Chrom	POS	REF	MUT	GENE	SUB	read_Proband	read_Parents	Filter	g(GnomAD_	train_AF	test_AF	ene_into	EA	SCORE	TRAIT
X	50635636	C	T	SHROOMR	146Q	1/1 (12/12)	0/15;10/24	PASS	3.6178857	0	0.015873016	40.98	77.23	44.63	3
X	49115764	G	A	GPLOW	T391I	1/1 (22/22)	0/14;15/33	PASS	3.5560114	0	0.015873016	78.96	46.48	39.13	0
X	66025129	C	T	VSIG4	G279E	1/1 (21/21)	0/19;8/22	PASS	1.5695154	0.0073	0.031746032	66.71	97.72	35.19	0
X	46879717	C	A	RP2	S334Y	1/1 (25/25)	0/14;20/33	PASS	3.4534758	0	0.015873016	94.75	22.48	25.13	0
X	129920470	G	A	UTP14A	V256M	1/1 (15/15)	0/13;16/27	PASS	1.7324997	0.0365	0.015873016	56.85	22.54	13.4	0
X	154712586	G	A	GAB3	P238S	1/1 (23/23)	0/17;16/30	PASS	1.3839017	0.0365	0.055555556	45.55	22.26	10.63	0
X	15287941	T	C	ASB11	S263G	1/1 (20/20)	0/18;11/25	PASS	1.7750994	0.0438	0.015873016	3.74	57.9	10.31	0
X	24363907	C	A	SUPT20H	Q383K	1/1 (27/27)	0/16;13/26	PASS	1.6481284	0.0292	0.031746032	23.53	21.24	9.29	0
X	20052928	T	C	MAP7D2	N182S	1/1 (26/26)	0/19;23/33	PASS	0.5764445	0.35036	0.214285714	40.17	44	8.95	0

RGP_TEST_22_PROBAND

Chrom	POS	REF	MUT	GENE	SUB	read_Proband	read_Parents	Filter	g(GnomAD_	train_AF	test_AF	ene_into	EA	SCORE	TRAIT
X	41170095	G	C	USP9X	G913R	1/1 (14/14)	./;15/33	PASS	6	0	0.015873016	98.86	44.77	58.3	4
X	155900572	A	T	VAMP7	M140L	0/1 (20/31)	./;24/33	PASS	4.7834884	0	0.015873016	49.21	68.54	52.77	0
X	9029354	C	T	FAM9B	E116K	1/1 (15/15)	./;16/31	PASS	5.2530274	0	0.015873016	54.61	39.17	40.05	0
X	139613237	C	T	MCF2	A524T	1/1 (16/16)	./;19/32	PASS	5.0220036	0	0.015873016	74.54	15.17	22.91	0
X	131282506	A	C	IGSF1	L395R	1/1 (20/20)	./;22/34	PASS	2.0427362	0.0146	0.031746032	68.35	29.65	19.1	0
X	47447773	G	A	ZNF41	A666V	1/1 (14/14)	./;21/37	PASS	2.5311079	0	0.015873016	37.02	17.83	12.79	0
Y	5100327	G	T	PCDH11Y	V917F	1/1 (13/13)	./;./	VQSR	0.2925721	0.48837	0.431818182	82.31	44.21	0	0
Y	9360389	C	T	TSPY8	S304F	0/1 (3/13)	./;./	Excess	6	0.65116	0.704545455	-	42.58	0	0
X	48195304	C	G	SSX5	E19Q	1/1 (15/15)	./;26/26	PASS	0.0188766	0.9781	0.992063492	67.09	39.31	0	0
X	120253978	C	CTGA	ZBTB33	indel	1/1 (12/12)	./;20/20	PASS	0.2169079	0.69343	0.73015873	84.41	0.001	0	0
X	83861196	G	C	CYLC1	R5T	1/1 (18/18)	./;25/38	PASS	1.7509362	0	0.015873016	51.36	66	0	0
X	137031305	C	A	GPR101	V124L	1/1 (13/13)	./;16/27	PASS	0.4367765	0.49635	0.507936508	26.98	25.46	0	1

HYPOTHESIS 3:

THERE ARE TWO RECESSIVE CAUSAL VARIANTS

Recessive variants: two inherited variants, one from each parent

Pathogenic:

SCORE_{recessive}

- *large Evolutionary Action scores*
- *rare or unseen in population*
- *true call*
- *gene is intolerant to variants*

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



PASS, DP, AD

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



Trait associations:

TRAIT

- *HPO*
- *ClinVar*
- *DisGeNet*
- *HumSavar*



HYPOTHESIS 3:

THERE ARE TWO RECESSIVE CAUSAL VARIANTS

Example :

RGP_TEST_10_PROBAND

Chrom	POS	REF	ALT	GENE	SUB	Probant	Father	Mother	Filter	(GnomAD	train_AF	test_AF	Gene_intol	EA	SCORE	TRAIT
1	1071926	G	A	RNF223	P214L	0/1 (24/49)	0/1 (18/36)	0/0 (0/35)	PASS	2.44933	0	0.011765	77.79	81.42	36.89	0
1	1072026	G	A		R181W	0/1 (31/51)	0/0 (0/39)	0/1 (32/47)	PASS	3.077605	0	0.011765		37.86		
14	55183707	T	G	DLGAP5	Q95H	0/1 (26/43)	0/0 (0/35)	0/1 (12/23)	PASS	4.907318	0	0.011765	34.62	52.6	31.27	0
14	55169541	G	A		T469I	0/1 (23/43)	0/1 (19/30)	0/0 (0/28)	PASS	1.399599	0.038889	0.023529		88		
16	88437287	C	T	ZNF469	P3245S	0/1 (21/34)	0/1 (24/41)	0/0 (0/38)	PASS	6	0	0.011765	35.98	52.78	30.83	4
16	88437831	G	A		R3426Q	0/1 (21/41)	0/0 (0/40)	0/1 (26/52)	PASS	2.07852	0.016667	0.017647		44.55		
14	19713015	T	C	OR11H2	N301S	0/1 (11/14)	0/0 (0/11)	0/1 (9/17)	PASS	2.170055	0.022222	0.017647	35.1	77.88	30.31	0
14	19713343	C	T		V192I	0/1 (9/19)	0/1 (18/42)	0/0 (0/19)	ExcessHet	0.514726	0.305556	0.382353		21.51		
11	1233177	G	A	MUC5B	G744S	0/1 (17/44)	0/1 (26/52)	0/0 (0/42)	PASS	2.926241	0.005556	0.011765	46.59	92.21	30.25	0
11	1250937	C	T		T4686M	0/1 (15/34)	0/0 (0/37)	0/1 (20/36)	PASS	1.168304	0.077778	0.052941		67.83		
22	22895680	A	G	IGLL5	T211A	1/1 (44/44)	0/1 (15/36)	0/1 (22/42)	PASS	1.652799	0.005556	0.041176	92.59	62.08	30.03	0



Causality hypotheses:

- de novo (one)
- x-linked (one)
- Inherited (two)

probabilities
of causality

Evolutionary Action $\Delta\varphi \approx \frac{\partial f}{\partial \gamma_i} \cdot \Delta\gamma_i$

Allele frequency

Variant call quality

Trait associations

gnomAD



mode



Model 3

[illegible]