

Project 1.3

Assessing Deleteriousness of SNVs in Carl's Genome

Yale University
CBB 752 Spring 2017

Linked SNPs
outside of gene

no effect on
protein production
or function

Causative SNPs
in gene

Non-coding SNP:

- changes amount of protein produced

Coding SNP:

- changes amino acid sequence

Protein

Regulatory sequences

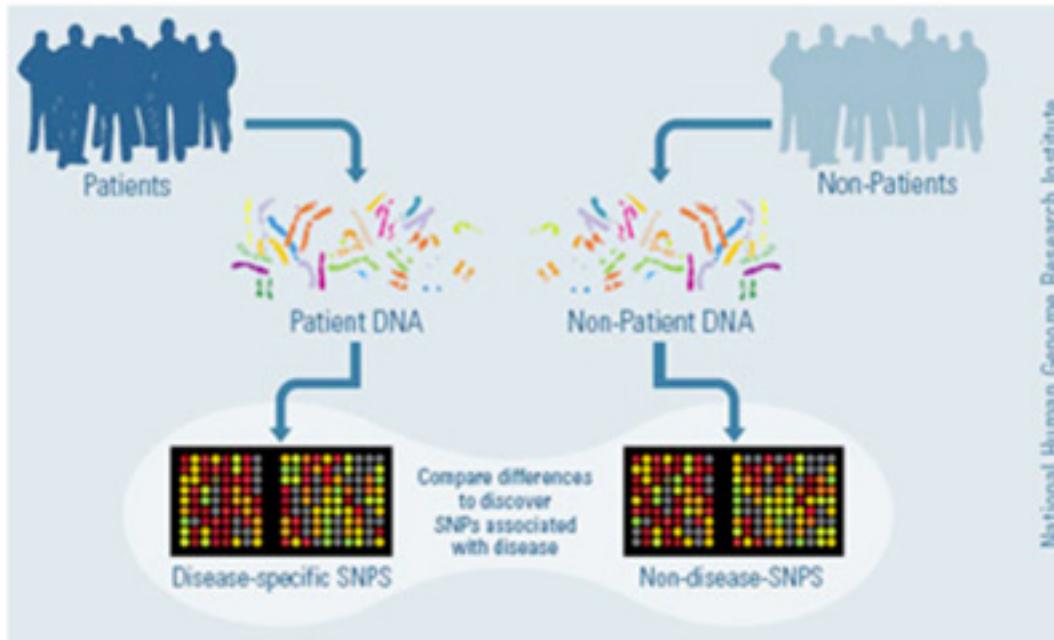
Gene

Coding region

What is Variant Prioritization?

The process of identifying deleterious SNVs

- SNP = seen in > 1% population
- SNV = no limitation on frequency



Principles of Variant Prioritization

Sequence:

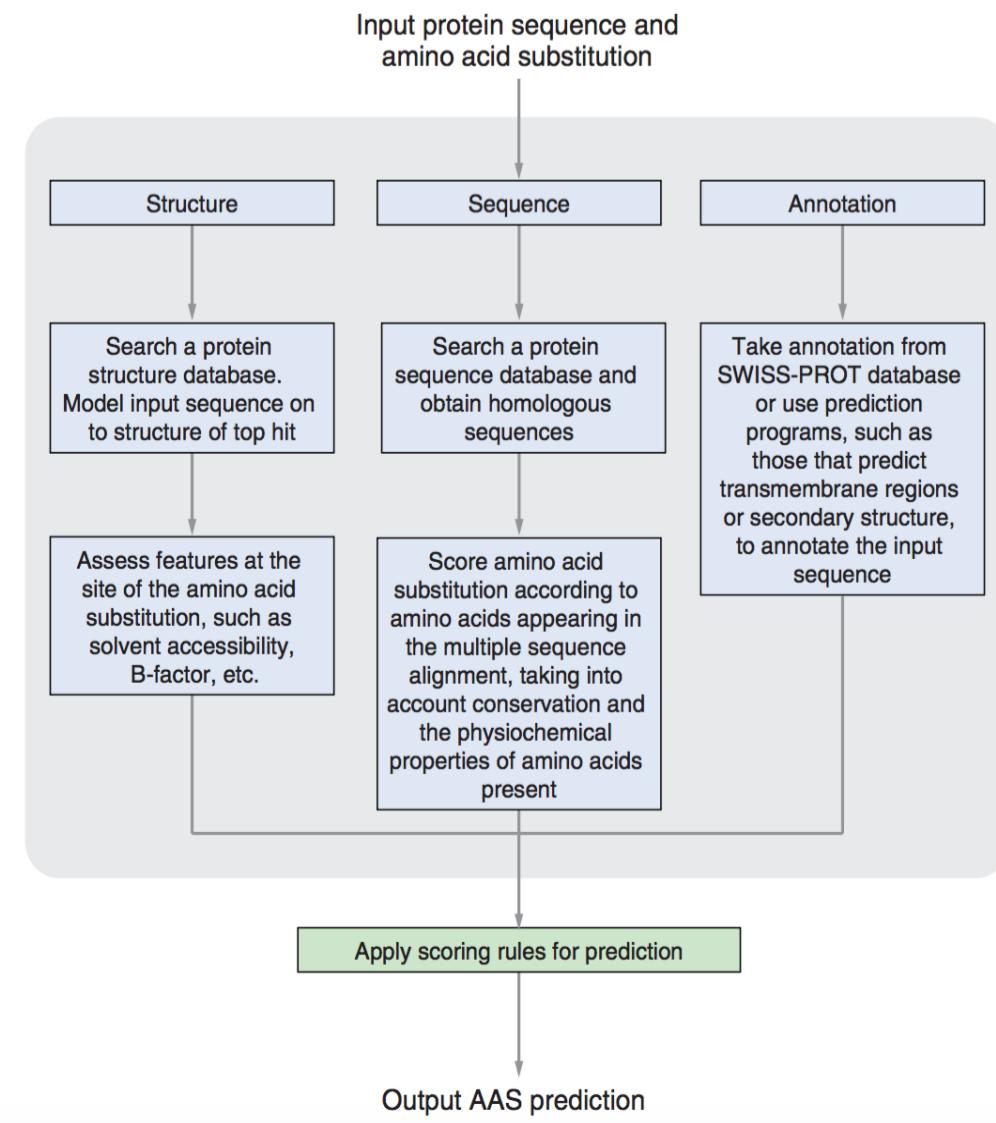
- synonymous or not?
- missense or nonsense?
- aa charge? size?

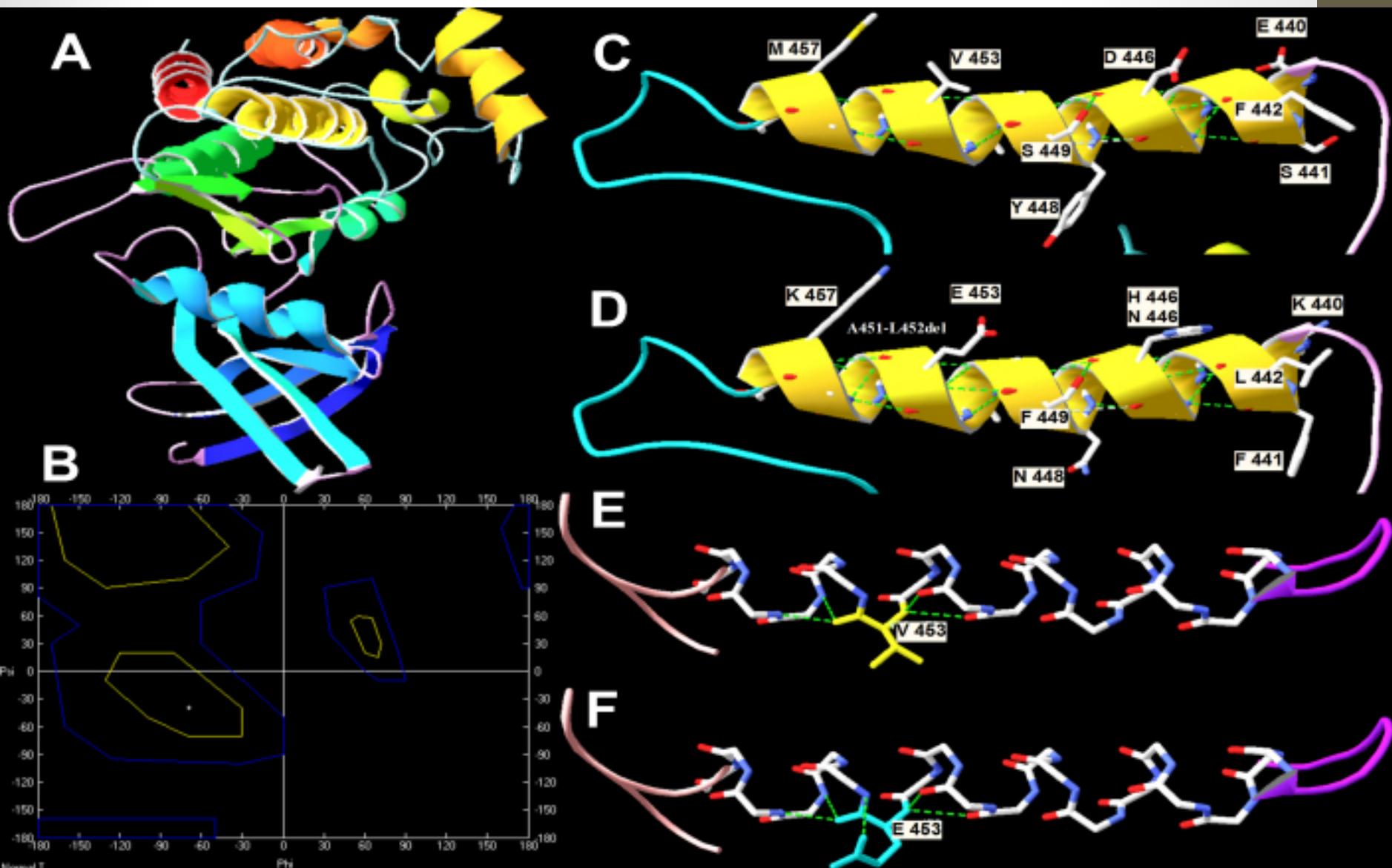
Structure:

- Phyre2, TMHMM

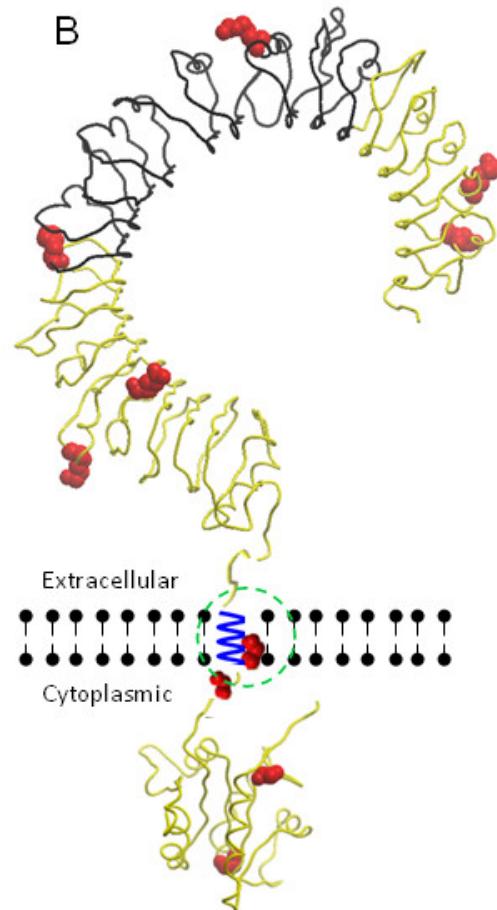
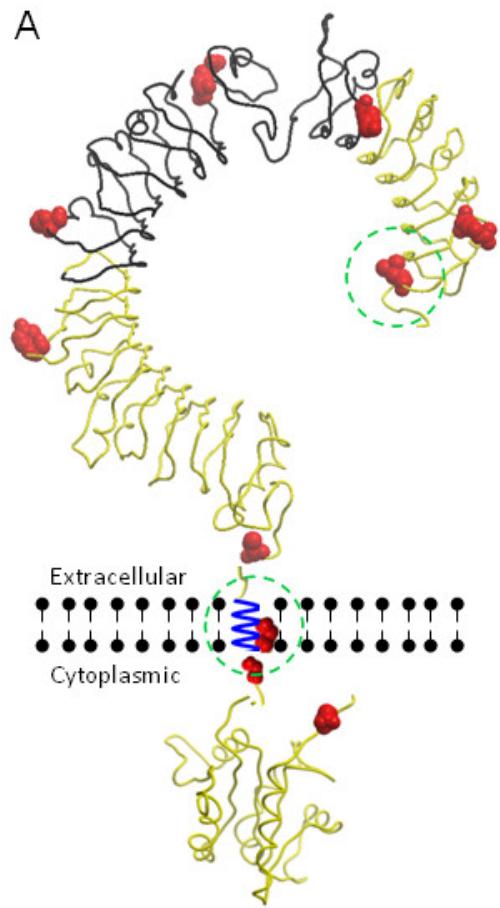
Annotation:

- dsSNP? GO term?

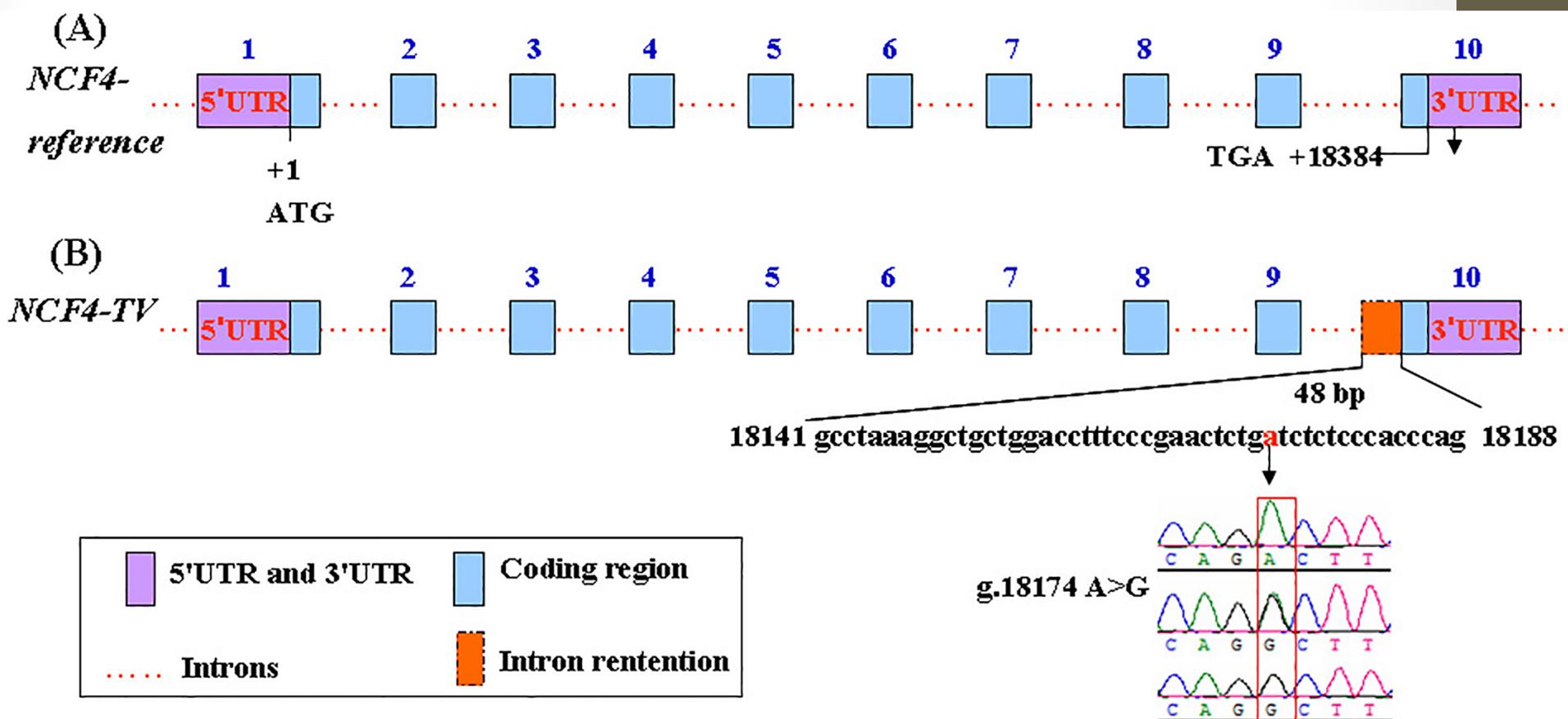


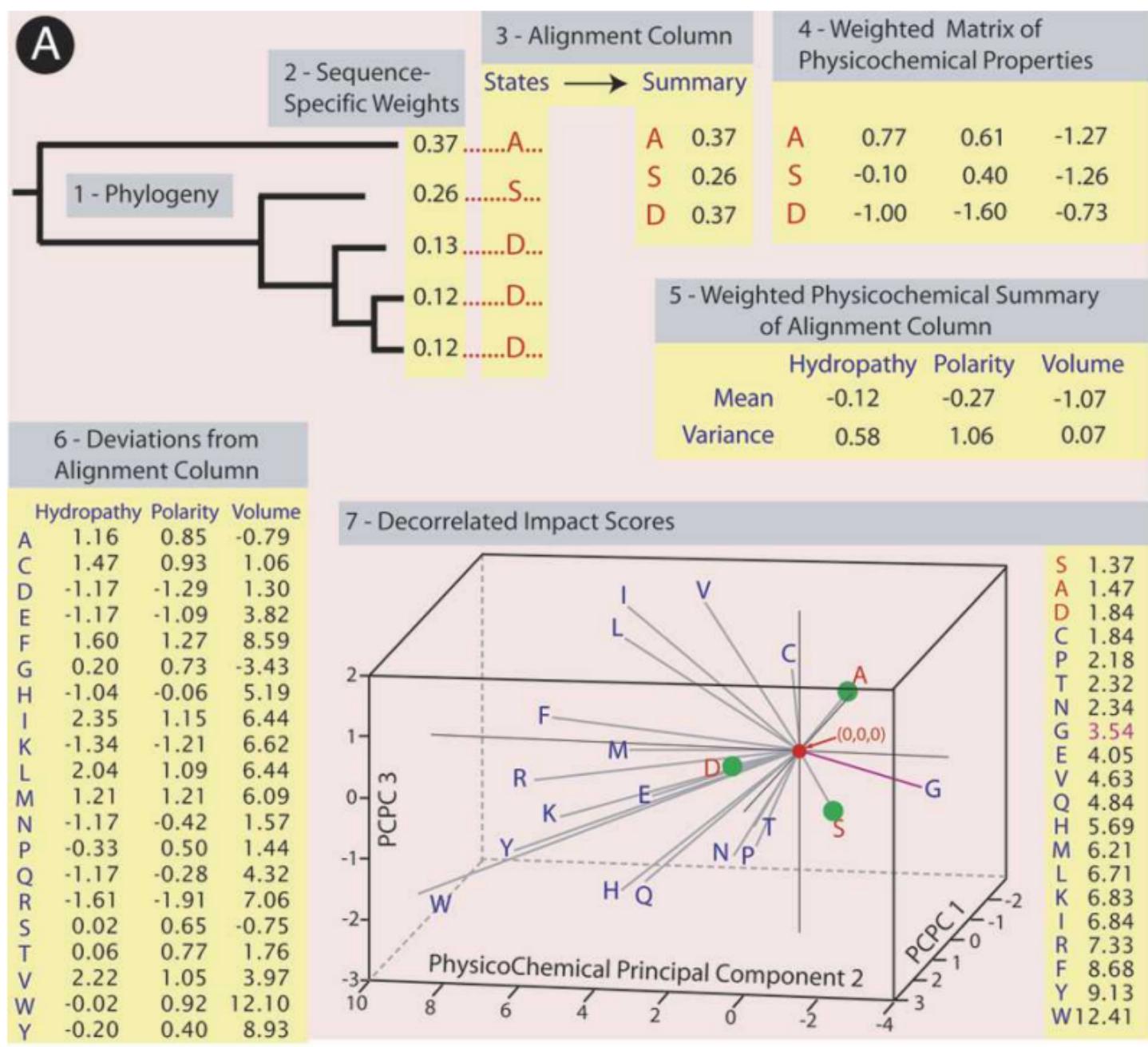


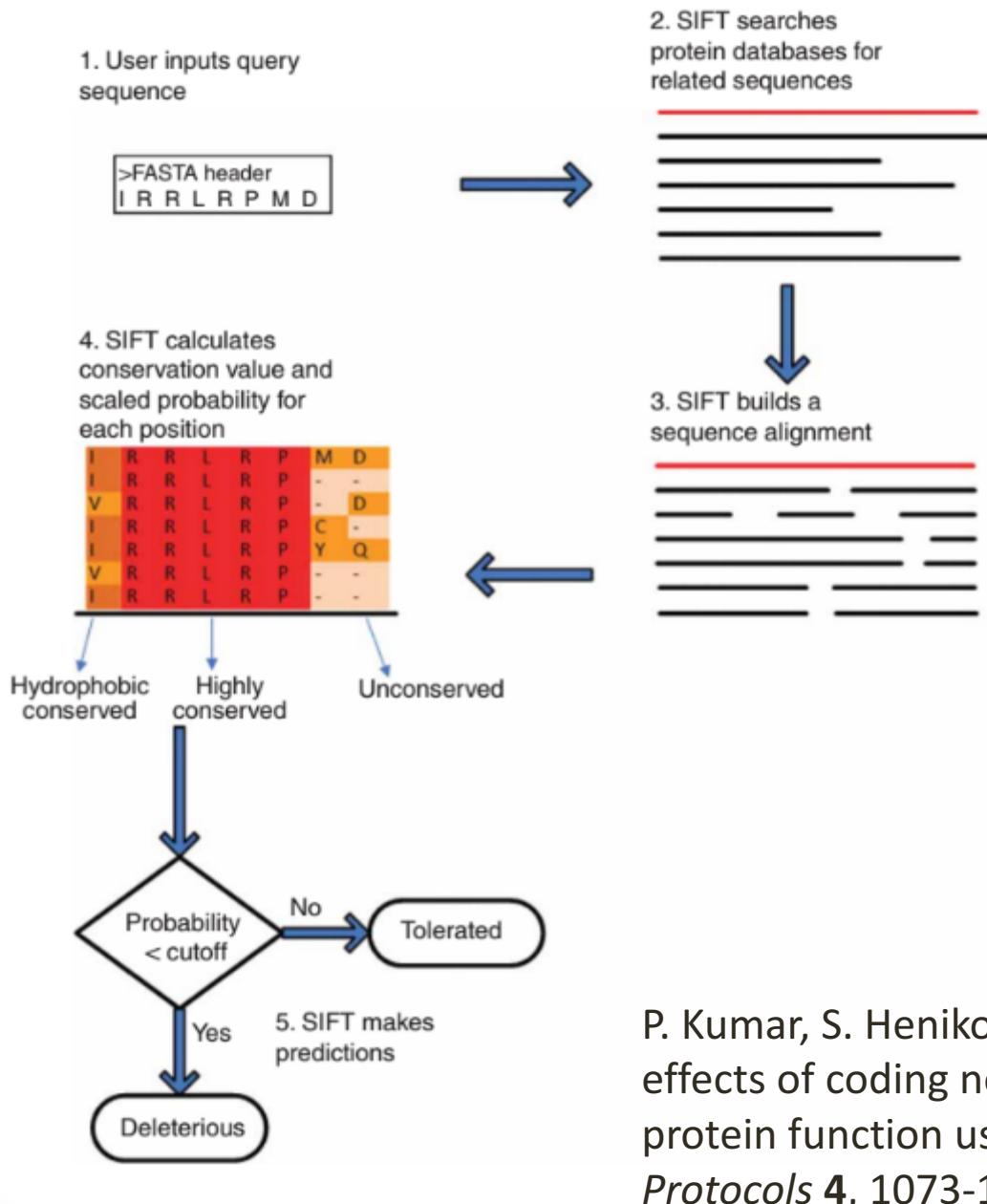
1. L. Zhang, L.-G. Gao, M. Zhang, X.-L. Zhou, Genotype-phenotype analysis of F-helix mutations at the kinase domain of TGFBR2, including a type 2 Marfan syndrome familial study. *Molecular Vision* **18**, 55-63 (2012).



S. A. Smith *et al.*, Adaptive evolution of Toll-like receptor 5 in domesticated mammals. *BMC Evolutionary Biology* **12**, 122 (2012).



A



P. Kumar, S. Henikoff, P. C. Ng, Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat. Protocols* 4, 1073-1081 (2009).

A

Training classifiers

SKEMPI Dataset

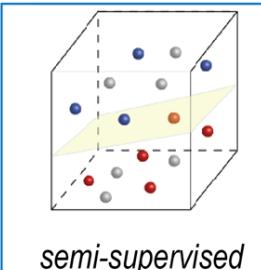
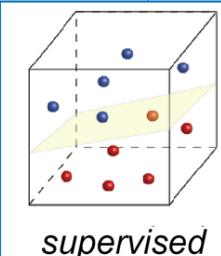
Random nsSNPs

Labeled nsSNPs



Unlabeled nsSNPs

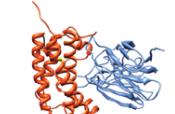
Unknown



B

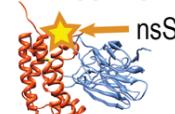
Calculating features

Wild-type PPI complex



FoldX modeling

Mutant PPI complex

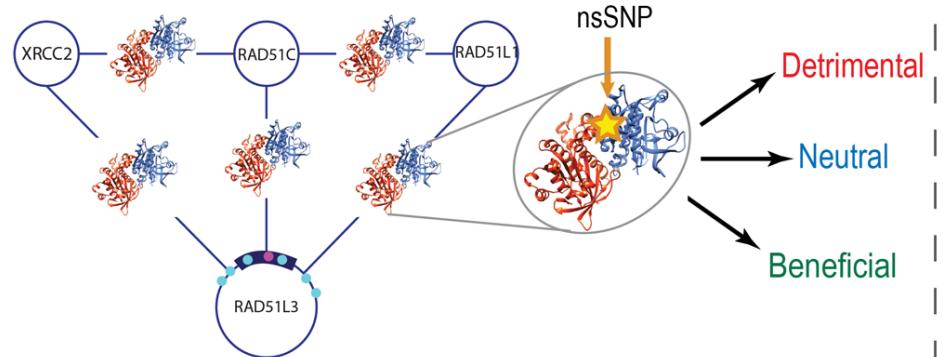


Binding energy terms



C

Predicting nsSNP effects on PPIs

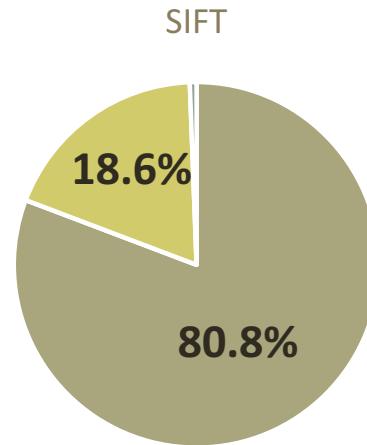
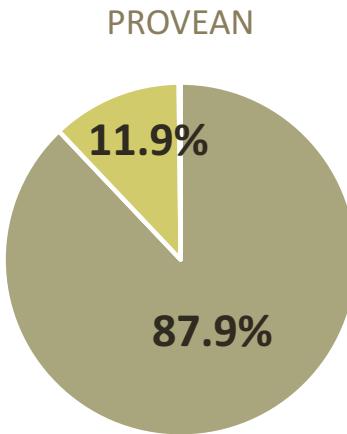
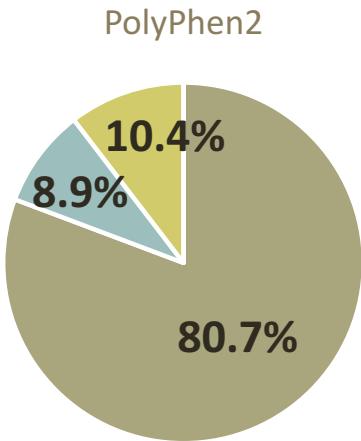


Programs to Classify SNPs

- Polyphen2
- PROVEAN
- SIFT



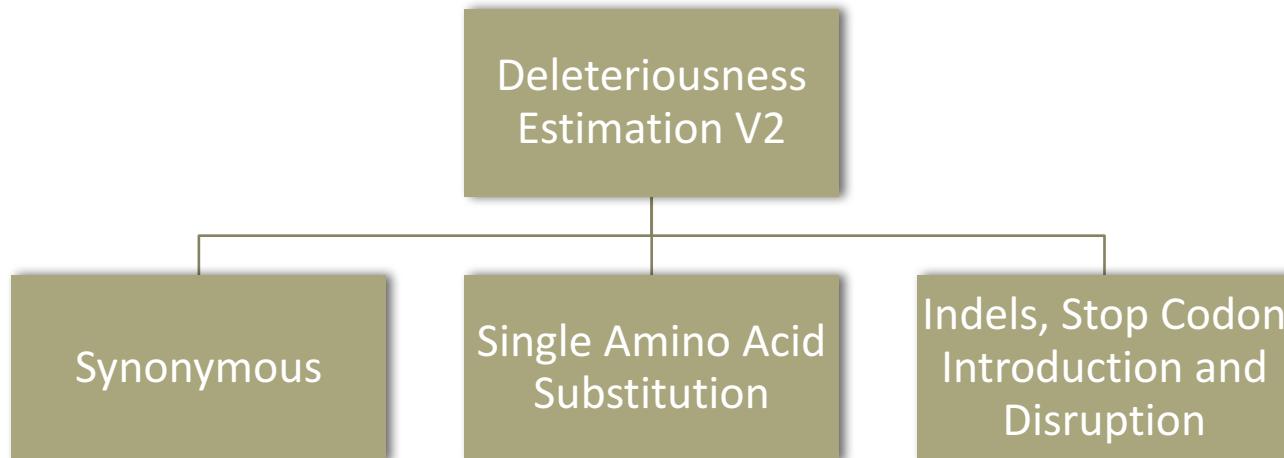
Comparison of Various Programs



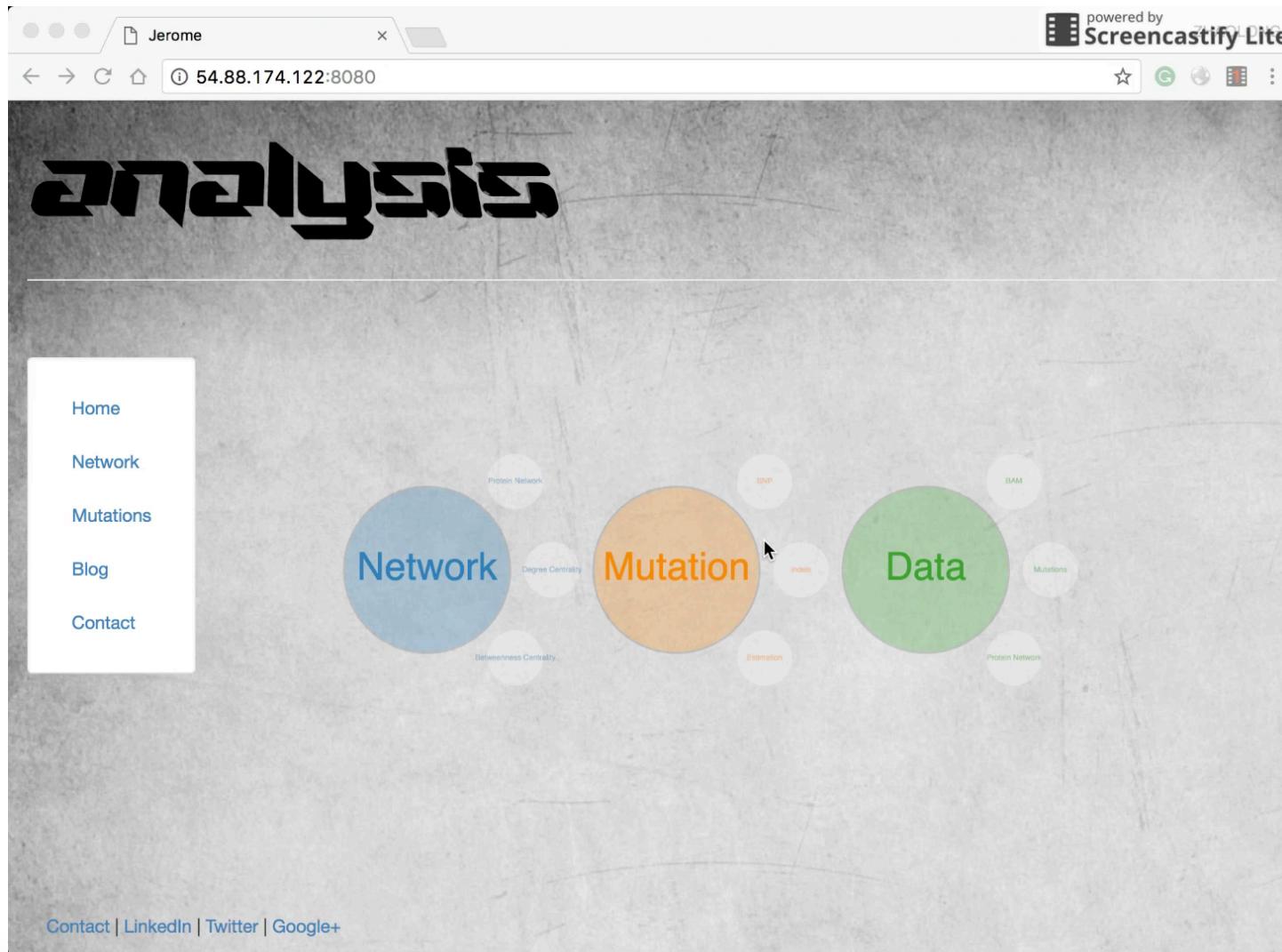
Brown = Neutral
Yellow = Damaging
Blue = Possibly damaging
Gray = Undetermined

Estimation Tool

- Deleteriousness Estimation V2
 - Estimating the deleteriousness of a mutation based on properties of amino acids and evolutionary constraint
 - Command line version and online version

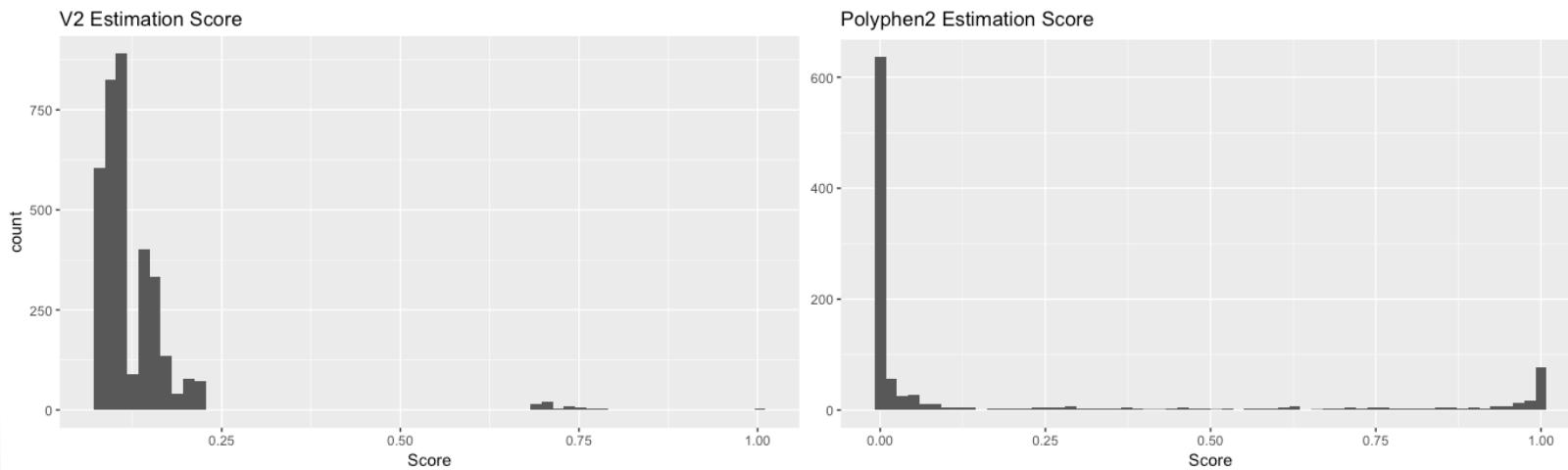


Website



Score

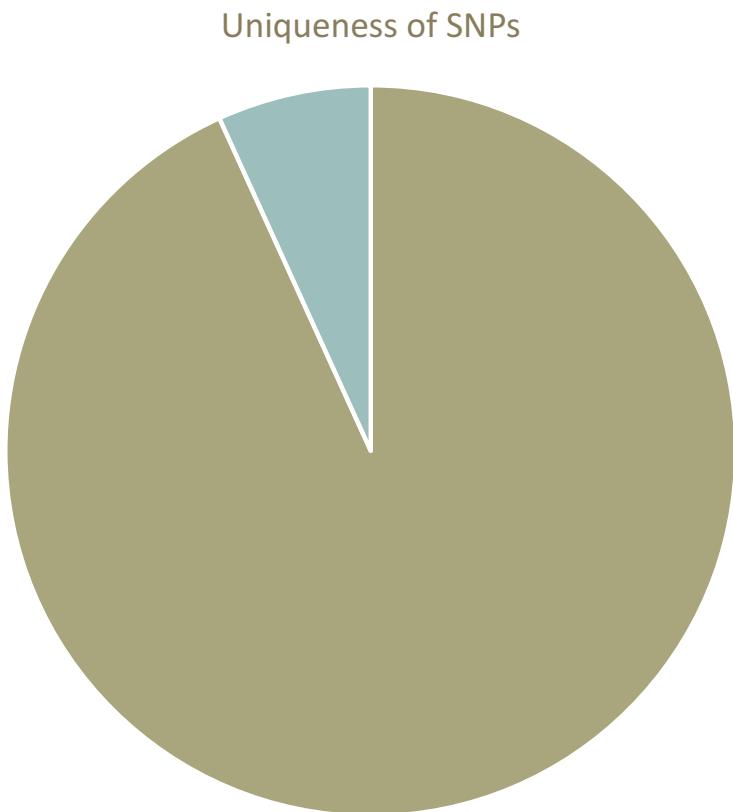
- Results
 - The distributions are basically the same
 - Our method provides more detailed information about the deleteriousness of some less deleterious mutations



Acknowledgements

- Mengting and Paul
- All the Professors in CBB752
- Carl

Analyzing the Zimmer Genome



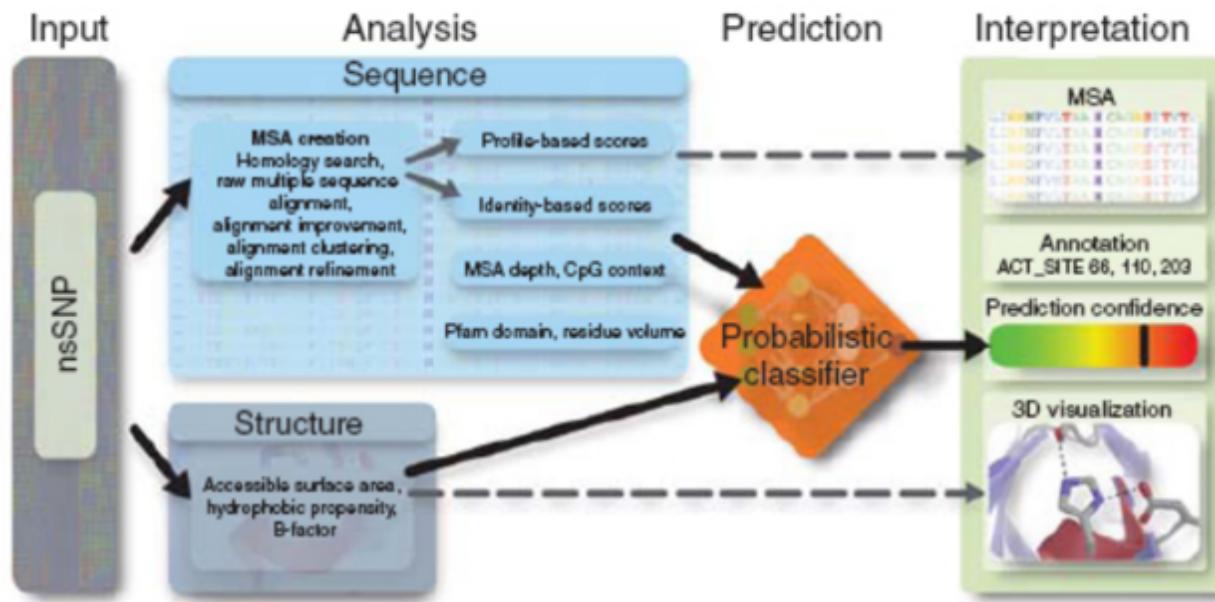
Brown = Present in dbSNP
Blue = Unique

Comparison of Various Programs

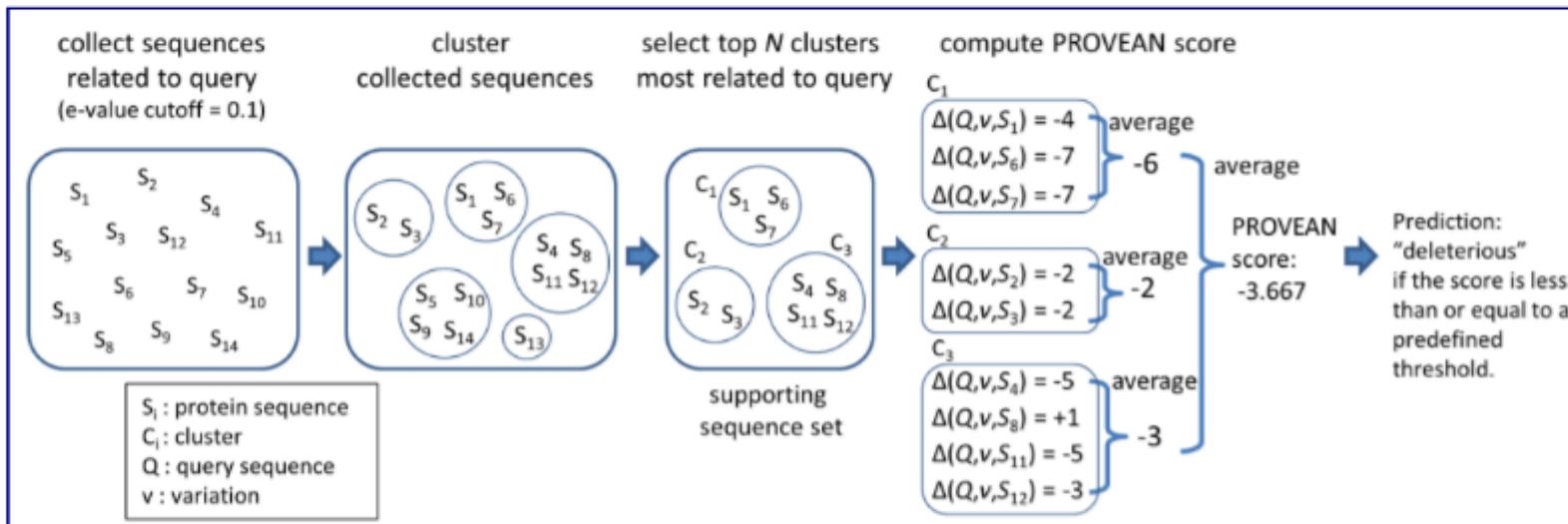
	PolyPhen2	PROVEAN	SIFT
Benign	2669	3104	2851
Possibly Damaging	293	0	0
Damaging	344	420	658
Undetermined	0	6	21

	PolyPhen2	PROVEAN	SIFT
Single Amino Acid Change	3281	3514	3514
Synonymous	15	10	10
Nonsense	10	6	6

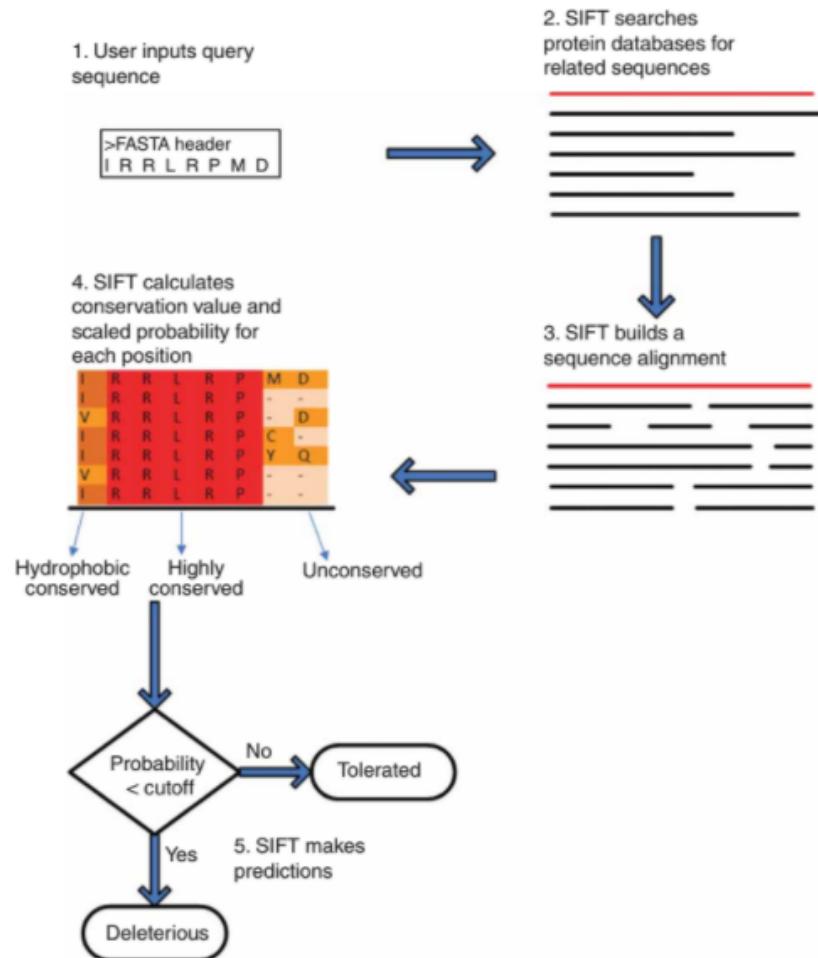
PolyPhen2



PROVEAN



SIFT



Polyphen2

Q9Y2Y4	0.936	# chr19:44156472 TC uc002oxf.1- PLAUR NP_002650	174	R	S	rs2227278	Q9Y2Y4	174	R	S	benign	0
Q9Y2Y4	1	# chr19:36206050 AC uc002oay.2+ ZBTB32 NP_055198	174	R	S	rs2227278	Q9Y2Y4	174	R	S	benign	0
Q9Y2Y4	1	# chr19:36206050 AC uc002oay.2+ ZBTB32 NP_055198	174	R	S	rs2227278	Q9Y2Y4	174	R	S	benign	0
P38398	0.911	# chr17:41223094 TC uc002ict.2- BRCA1 NP_009231	1613	S	G	rs1799966	P38398	1613	S	G	benign	0.255
P38398	0.911	# chr17:41223094 TC uc002ict.2- BRCA1 NP_009231	1613	S	G	rs1799966	P38398	1613	S	G	benign	0.255
P38398	0.911	# chr17:41223094 TC uc002ict.2- BRCA1 NP_009231	1613	S	G	rs1799966	P38398	1613	S	G	benign	0.255
P38398	0.911	# chr17:41223094 TC uc002ict.2- BRCA1 NP_009231	1613	S	G	rs1799966	P38398	1613	S	G	benign	0.255
P38398	0.911	# chr17:41223094 TC uc002ict.2- BRCA1 NP_009231	1613	S	G	rs1799966	P38398	1613	S	G	benign	0.255
P38398	0.911	# chr17:41223094 TC uc002ict.2- BRCA1 NP_009231	1183	K	R	rs16942	P38398	1183	K	R	benign	0
P38398	1	# chr17:41244400 TC uc002ict.2- BRCA1 NP_009231	1038	E	G	rs16941	P38398	1038	E	G	possibly damaging	0.936
P38398	0.9568	# chr17:41244435 TC uc002ict.2- BRCA1 NP_009231	871	P	L	rs799917	P38398	871	P	L	benign	0
P38398	1	# chr17:41244936 GA uc002ict.2- BRCA1 NP_009231	693	D	N	rs4986650	P38398	693	D	N	benign	0
P38398	1	# chr17:41245471 CT uc002ict.2- BRCA1 NP_009231	356	Q	R	rs1799950	P38398	356	Q	R	probably damaging	0.998
Q14181	0.273	# chr17:41246481 TC uc002ict.2- BRCA1 NP_009231	583	G	R	rs487989	Q14181	583	G	R	possibly damaging	0.474
Q14181	0.0987	# chr11:65064690 GA uc001odj.2+ POLA2 NP_002680	58	M	V	rs1847626	Q6PML9	58	M	V	benign	0
Q6PML9	1	# chr4:42003671 AG uc003gw1.2+ SLC30A9 NP_006336	97	T	A	rs2581423	Q6PML9	97	T	A	benign	0
Q6PML9	1	# chr4:42020142 AG uc003gw1.2+ SLC30A9 NP_006336	235	R	C	rs11550929	Q8IVD9	235	R	C	benign	0.041
Q8IVD9	0.172	# chr7:44444122 GA uc003tkz.2- NUDCD3 NP_056147	3	T	P	rs307807	Q8IVD9	3	T	P	benign	0
Q8IVD9	1	# chr7:44530193 TG uc003tkz.2- NUDCD3 NP_056147	588	F	P	?	Q8IVD9	588	F	P	benign	0.028

- Probability calculated through Bayes classifier
 - Benign (0-.45), possibly damaging (.45-.95), probably damaging (.95-1)

PROVEAN & SIFT

34	2,17984421,T,G	ENSP000000841784	33428	-1	UUU [A/C]CC CII	2868	I	P	Single AA Change	-0.07	Deletion(s)	65	30	0.118	Tolerated
35	2,17984421,T,G	ENSP000000841784	33428	-1	UUU [A/C]CC CII	2868	I	P	Single AA Change	-0.07	Deletion(s)	65	30	0.118	Tolerated
36	2,17984421,T,G	ENSP000000841784	33428	-1	UUU [A/C]CC CII	2868	I	P	Single AA Change	-0.07	Deletion(s)	65	30	0.118	Tolerated
37	2,17984421,T,G	ENSP000000841784	33428	-1	UUU [A/C]CC CII	2868	I	P	Single AA Change	-0.07	Deletion(s)	65	30	0.118	Tolerated
38	2,17984421,T,G	ENSP000000841784	33428	-1	UUU [A/C]CC CII	2868	I	P	Single AA Change	-0.07	Deletion(s)	65	30	0.118	Tolerated
39	2,17984421,T,G	ENSP000000841784	33428	-1	UUU [A/C]CC CII	2868	I	P	Single AA Change	-0.07	Deletion(s)	65	30	0.118	Tolerated
40	2,17984421,T,G	ENSP000000841784	33428	-1	UUU [A/C]CC CII	2868	I	P	Single AA Change	-0.07	Deletion(s)	65	30	0.118	Tolerated
41	2,17984421,T,G	ENSP000000841784	33428	-1	UUU [A/C]CC CII	2868	I	P	Single AA Change	-0.07	Deletion(s)	65	30	0.118	Tolerated
42	2,17984421,T,G	ENSP000000841784	33428	-1	UUU [A/C]CC CII	2868	I	P	Single AA Change	-0.07	Deletion(s)	65	30	0.118	Tolerated
43	2,17984421,T,G	ENSP000000841784	33428	-1	UUU [A/C]CC CII	2868	I	P	Single AA Change	-0.07	Deletion(s)	65	30	0.118	Tolerated
44	2,17984421,T,G	ENSP000000841784	33428	-1	UUU [A/C]CC CII	2868	I	P	Single AA Change	-0.07	Deletion(s)	65	30	0.118	Tolerated
45	2,17984421,T,G	ENSP000000841784	33428	-1	UUU [A/C]CC CII	2868	I	P	Single AA Change	-0.07	Deletion(s)	65	30	0.118	Tolerated
46	2,17984421,T,G	ENSP000000841784	33428	-1	UUU [A/C]CC CII	2868	I	P	Single AA Change	-0.07	Deletion(s)	65	30	0.118	Tolerated

Comparison of Various Programs

