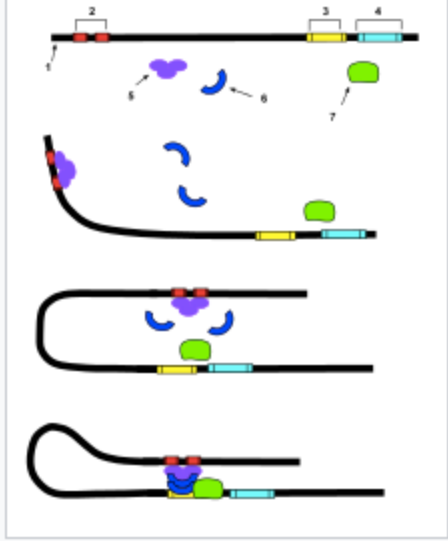
Team Bayes Pair

What's an enhancer ?

- Important for regulation
- Difficult to characterize
- 100-1500 bp
- Transcription factor binding
- May be located upstream or downstream of regulated gene



Seen here is a four step diagram depicting the usage of an enhancer. Within this DNA sequence, protein(s) known as transcription factor(s) bind to the enhancer and increase the activity of the promoter.

1. DNA
2. Enhancer
3. Promoter
4. Gene
5. Transcription Activator Protein
6. Mediator Protein
7. RNA Polymerase

Predicting functional effects of rare SNPS In enhancer regions.

Methods:

- Use 1000 genome project's SNP distributions to detect rare variants across patient's tumor genome.
- Run Deep Sea's Convolutional Neural Network
  - Trained on experimental enhancer data.

Jian Zhou, Olga G. Troyanskaya. Predicting the Effects of Noncoding Variants with Deep learning-based Sequence Model. Nature Methods (2015).

  - Area under the curve (AUC): 0.958

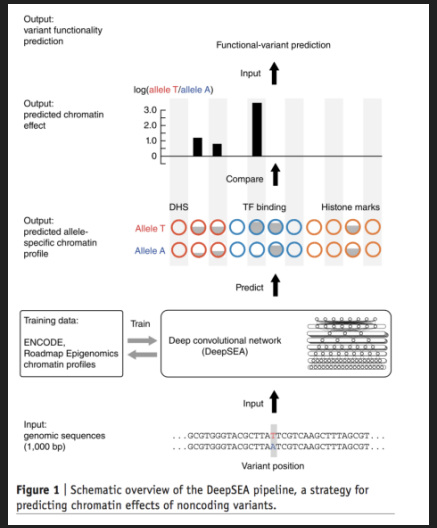
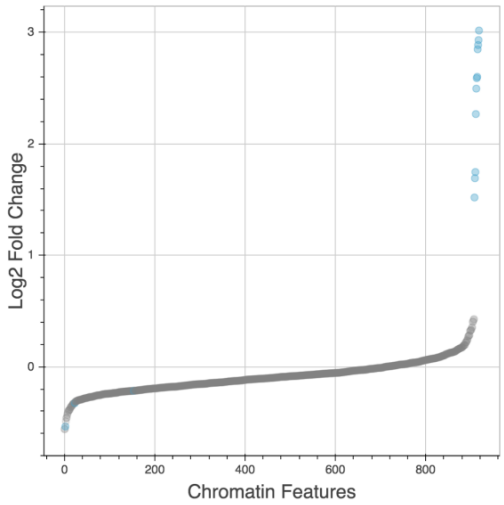


Figure 1 | Schematic overview of the DeepSEA pipeline, a strategy for predicting chromatin effects of noncoding variants.

Results

DeepSEA Home Sequence Profiler Help

Variant row number: 11 chr2:19775191 G>A rs1344649 Go to Genome Browser



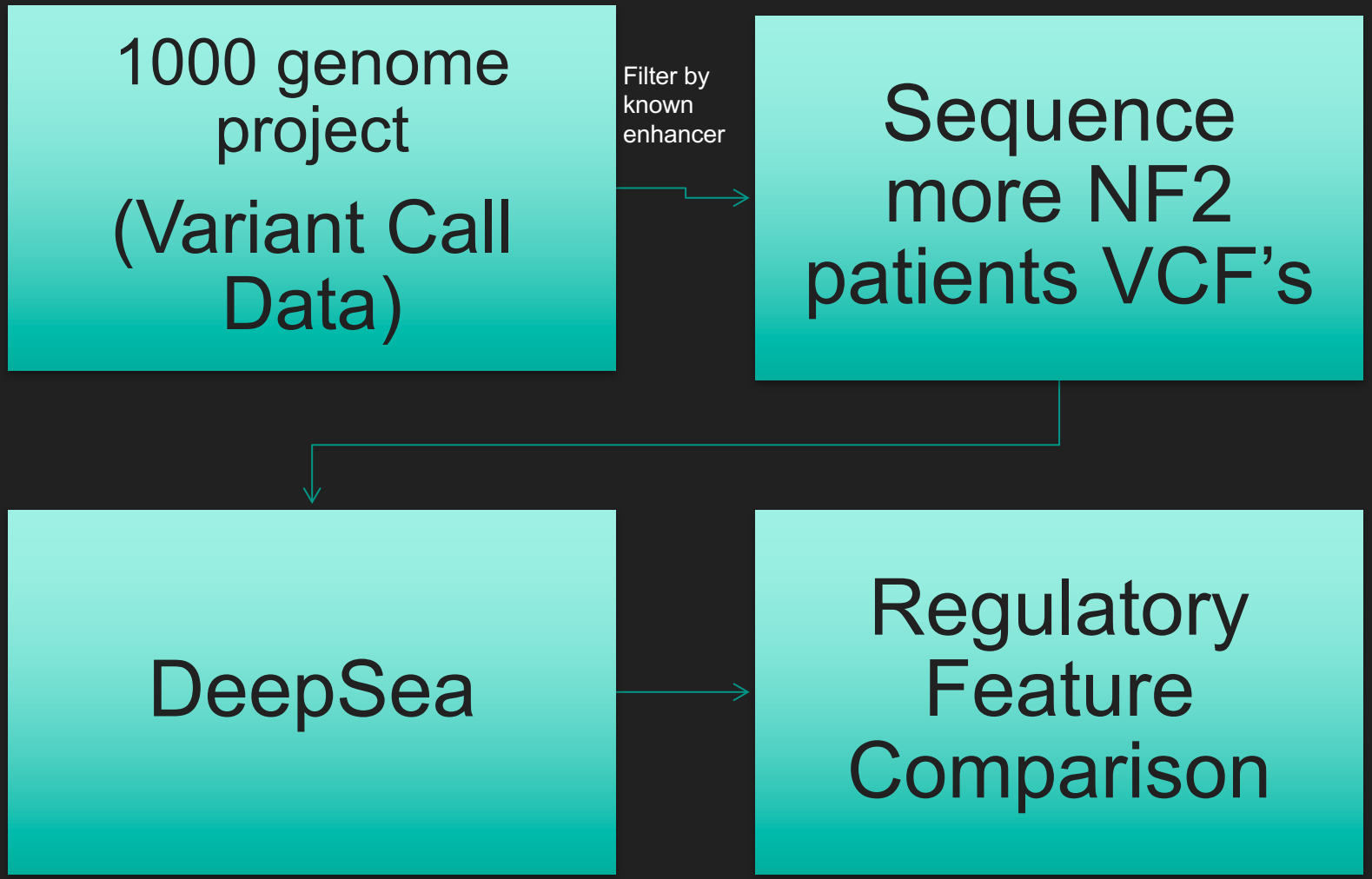
Functional significance score: 2.0745e-3

Chromati...	Cell type	Treat...	Effect (Lo...	*E-value*	P(Reference)	P(Alternative)	
MafK	**IMR90**	**None**	**2.8481**	**0.001760**	**0.008544**	**0.062269**	**(**
MafK	**HepG2**	**None**	**3.0162**	**0.002311**	**0.013320**	**0.105760**	**(**
MafF	**HepG2**	**None**	**2.8861**	**0.002561**	**0.005705**	**0.044044**	**(**
MafK	**HepG2**	**None**	**2.9305**	**0.002574**	**0.005394**	**0.043027**	**(**
MafK	**H1-hESC**	**None**	**2.4953**	**0.002774**	**0.000984**	**0.005987**	**(**
MafK	**HeLa-S3**	**None**	**2.267**	**0.003158**	**0.001808**	**0.009181**	**(**
MafF	**K562**	**None**	**2.589**	**0.003559**	**0.002366**	**0.014896**	**(**
MafK	**K562**	**None**	**2.6009**	**0.003851**	**0.001341**	**0.008563**	**(**
Bach1	**K562**	**None**	**1.6912**	**0.007888**	**0.000222**	**0.000736**	**(**
Bach1	**H1-hESC**	**None**	**1.5184**	**0.008013**	**0.000457**	**0.001349**	**(**
NF-E2	**K562**	**None**	**1.747**	**0.012003**	**0.000122**	**0.000421**	**(**
TCF7L2	**PANC-1**	**None**	**-0.32612**	**0.016214**	**0.015112**	**0.012126**	**(**
KAP1	**HEK293**	**None**	**-0.21581**	**0.016294**	**0.058268**	**0.050702**	**(**
GATA3	**SH-SY5Y**	**None**	**-0.53603**	**0.019030**	**0.037895**	**0.026366**	**(**
ZNF274	HeLa-S3	None	0.16856	0.022171	0.000247	0.000278	(
NRSF	PFSK-1	None	-0.36801	0.023927	0.004565	0.003511	(
GATA2	SH-SY5Y	None	-0.56018	0.029641	0.035204	0.024033	(
Nrf1	HeLa-S3	None	0.34816	0.033650	0.000255	0.000325	(
Pol2(b)	HUVEC	None	-0.28643	0.033977	0.002762	0.002269	(
TAF1	PFSK-1	None	-0.24212	0.034975	0.002255	0.001909	(

- 236 Rare SNPs identified in tumor
- 35 SNPs with a functional significance score < 0.01

Future Directions

- Apply our process to other NF2 patient samples to further hone in on potentially biologically relevant SNPs in enhancer regions
- Validate findings in vitro
- Use identified enhancer regions as new drug targets
- Refine ML and DL algorithms to improve performance and extend utility


1000 genome project
(Variant Call Data)
Filter by known enhancer
Sequence more NF2 patients VCF's
DeepSea
Regulatory Feature Comparison

The Team!

- Maxwell Murphy
  - https://www.linkedin.com/in/maxwellrmurphy/
- Jordan Wilheim
  - https://www.linkedin.com/in/jordan-wilheim/
- Vince Kumar
  - https://www.linkedin.com/in/vince-kumar/
- Jim Wright