WHO ARE WE

The Interuniversity Institute of Bioinformatics in Brussels

Brussels, Belgium

Harness and scale bioinformatics expertise

Provide infrastructure

Provide scientific advancements

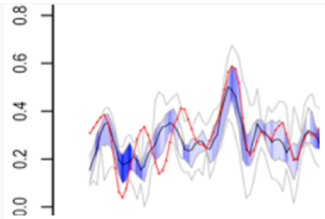
The directors

Tom Lenaerts

Wim Wranken

WHO ARE WE

15 ONGOING PROJECTS



PREDICTING BIOPHYSICAL CHARACTERISTICS OF PROTEINS FROM THEIR AMINO ACID SEQUENCE

Structural Bioinformatics

...FCG...

...FYG...

ASSESSING THE LIKELY EFFECT OF AMINO ACID MUTATIONS ON A PROTEIN (AND THE ORGANISM IT IS IN)

Structural Bioinformatics

DETERMINING SIMILARITY BETWEEN PROTEINS IN BIOPHYSICAL SPACE.

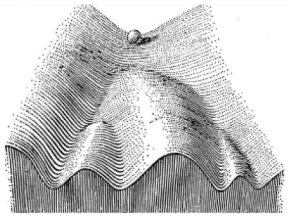
Structural Bioinformatics

DEVELOPING QUANTITATIVE MODELS FROM GENE REGULATORY NETWORKS TO MICROBIAL COMMUNITIES

Systems Biology

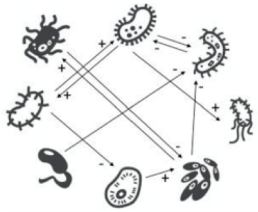
BRUSSELS INTELLIGENT ICT FOR GENOMIC HIGH THROUGHPUT ANALYSIS (BRIGHTANALYSIS)

Bioinformatics Applications, Genomics and Genetic Bioinformatics



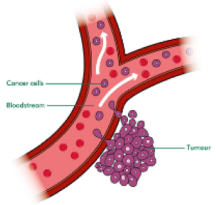
UNDERSTANDING HEMATOPOIESIS IN LIGHT OF LEUKEMIA AND OTHER HEMATOLOGICAL DISEASES

Systems Biology



MODELING THE DYNAMICS OF MICROBIAL COMMUNITIES

Systems Biology

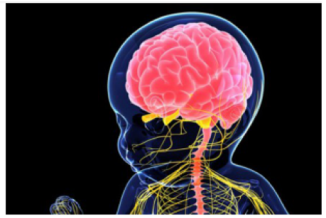


MATHEMATICS OF METASTATIC INEFFICIENCY (TÉLÉVIE PROJECT)

Systems Biology

FRIA ; CROSSING THE MONOGENIC BARRIER: DEVELOPMENT OF CLINICALLY COMPETENT METHODS FOR NEURODEVELOPMENTAL DISEASES

Genomics and Genetic Bioinformatics

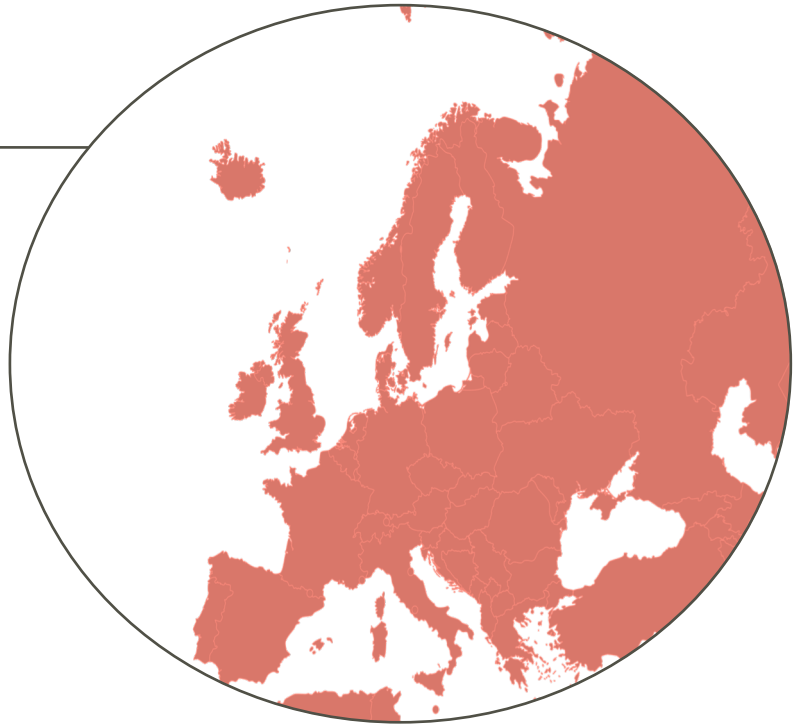


ARC : DECIPHERING THE GENETIC ARCHITECTURE FROM OLIGO- TO POYLYGENIC IN NEURODEVELOPMENTAL DISEASES.

Genomics and Genetic Bioinformatics

GENETIC DISEASES: SEEKING UNDERSTANDING

30 million people affected by a **rare disease** in Europe

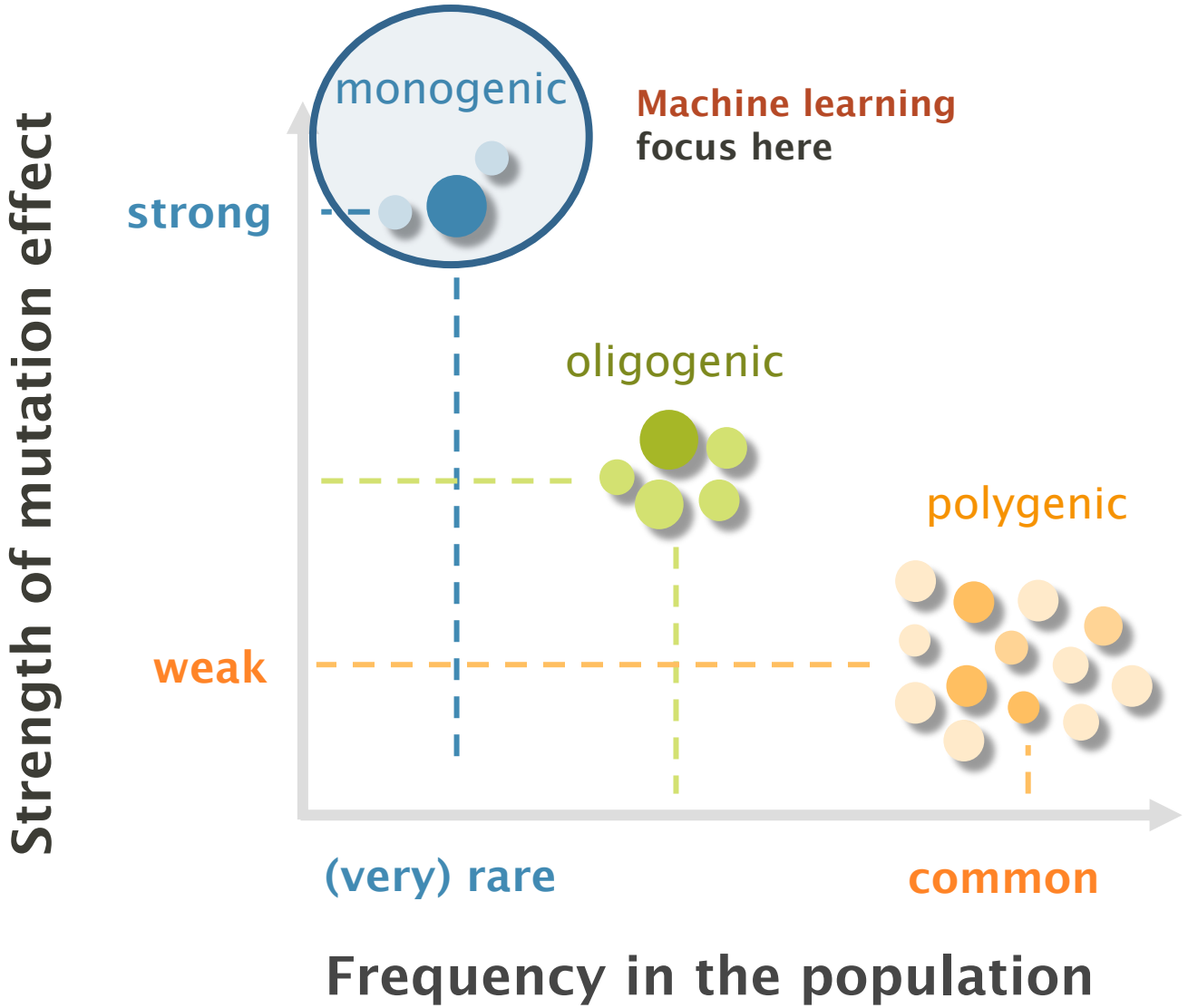


Many have **genetic origin**

Difficult to associate a **phenotype** with **genetic cause**

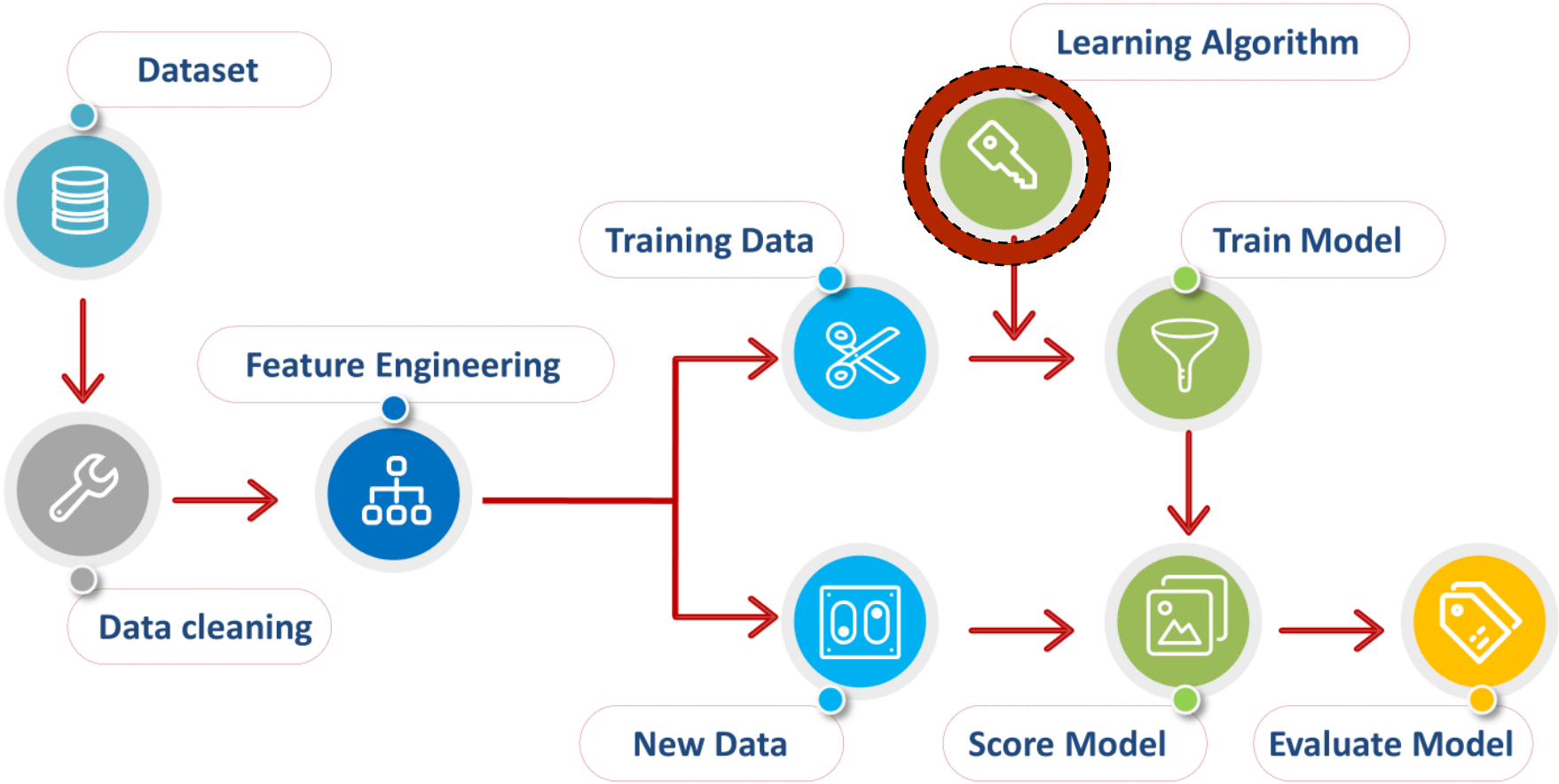
Source: EURORDIS

MACHINE LEARNING FOCUS ON MONOGENIC



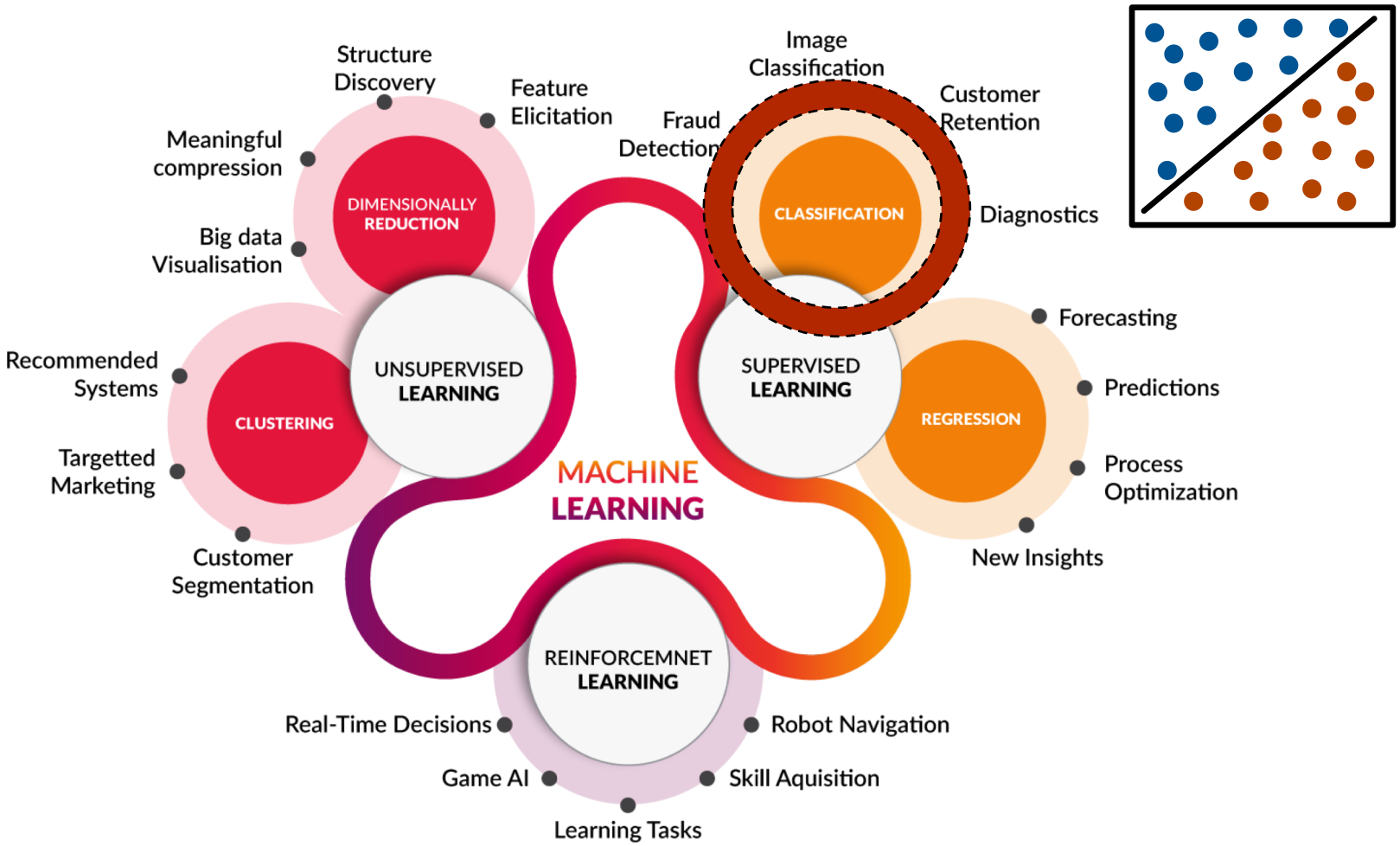
Source: McCarthy, *et al.* (2008). *Nature reviews genetics*. 9(5), 356.

THE PROCESS OF MACHINE LEARNING



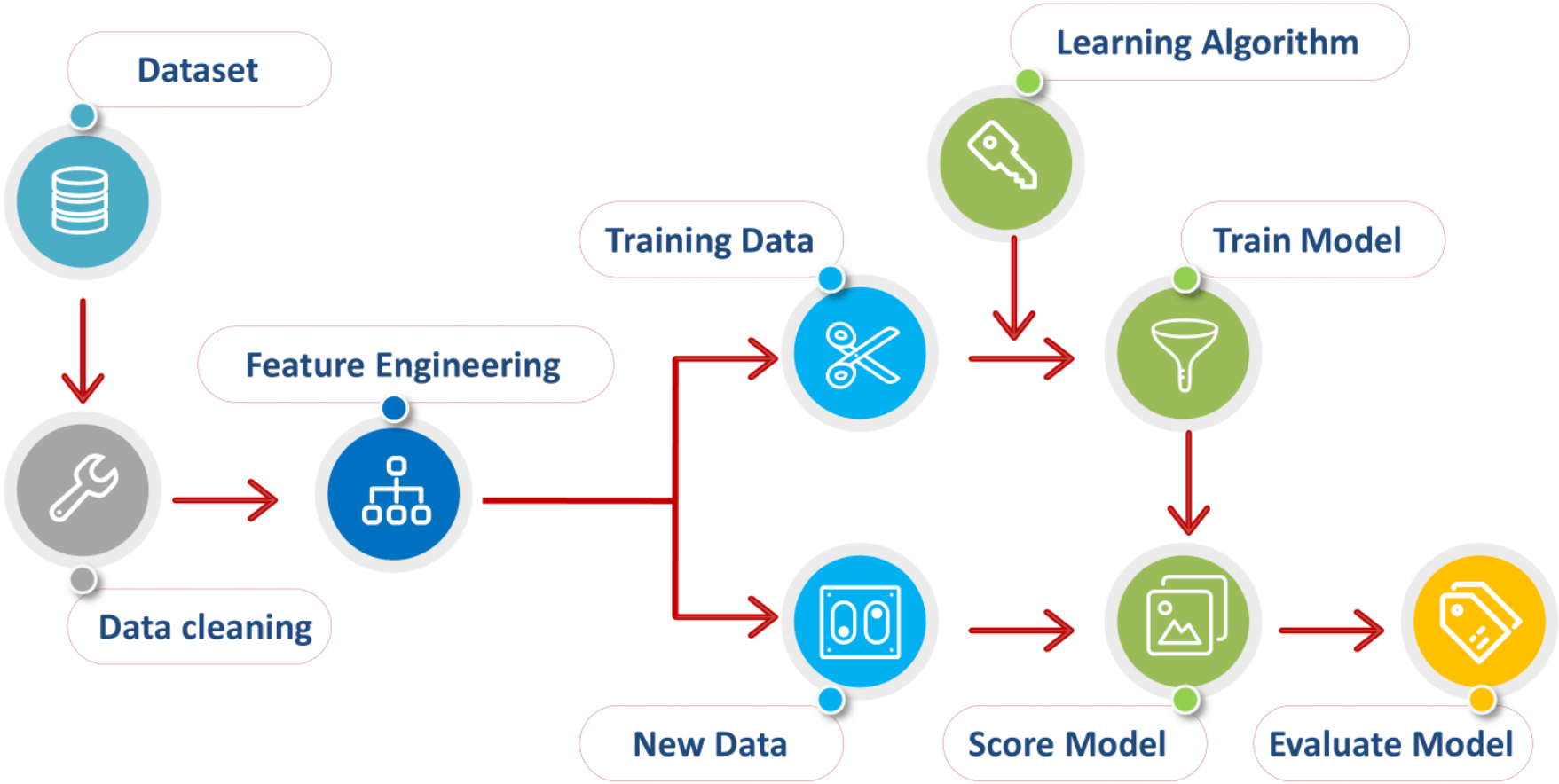
Source: https://towardsdatascience.com

DIFFERENT MACHINE LEARNING ALGORITHMS



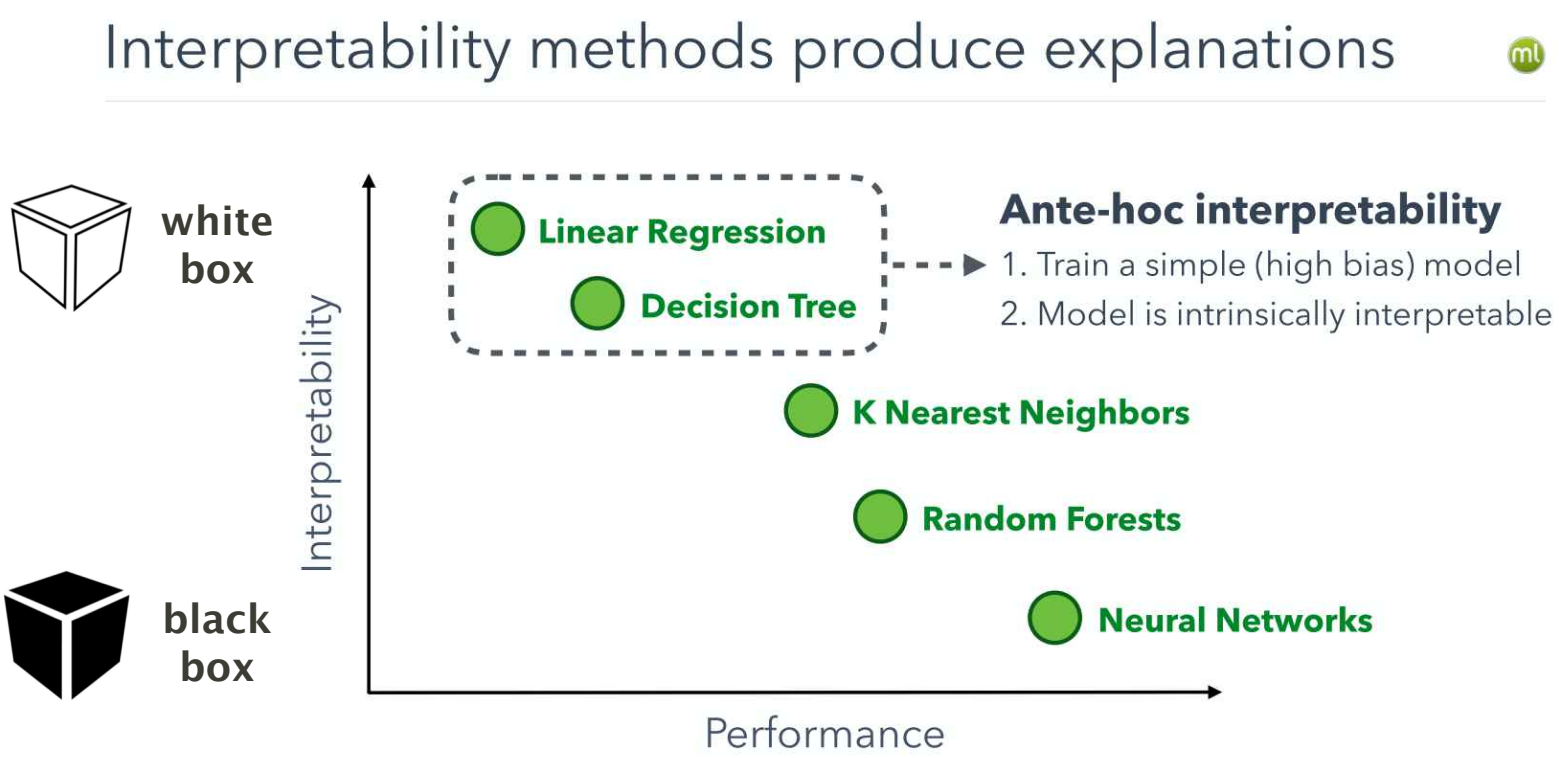
Source: https://towardsdatascience.com

THE PROCESS OF MACHINE LEARNING



Source: https://towardsdatascience.com

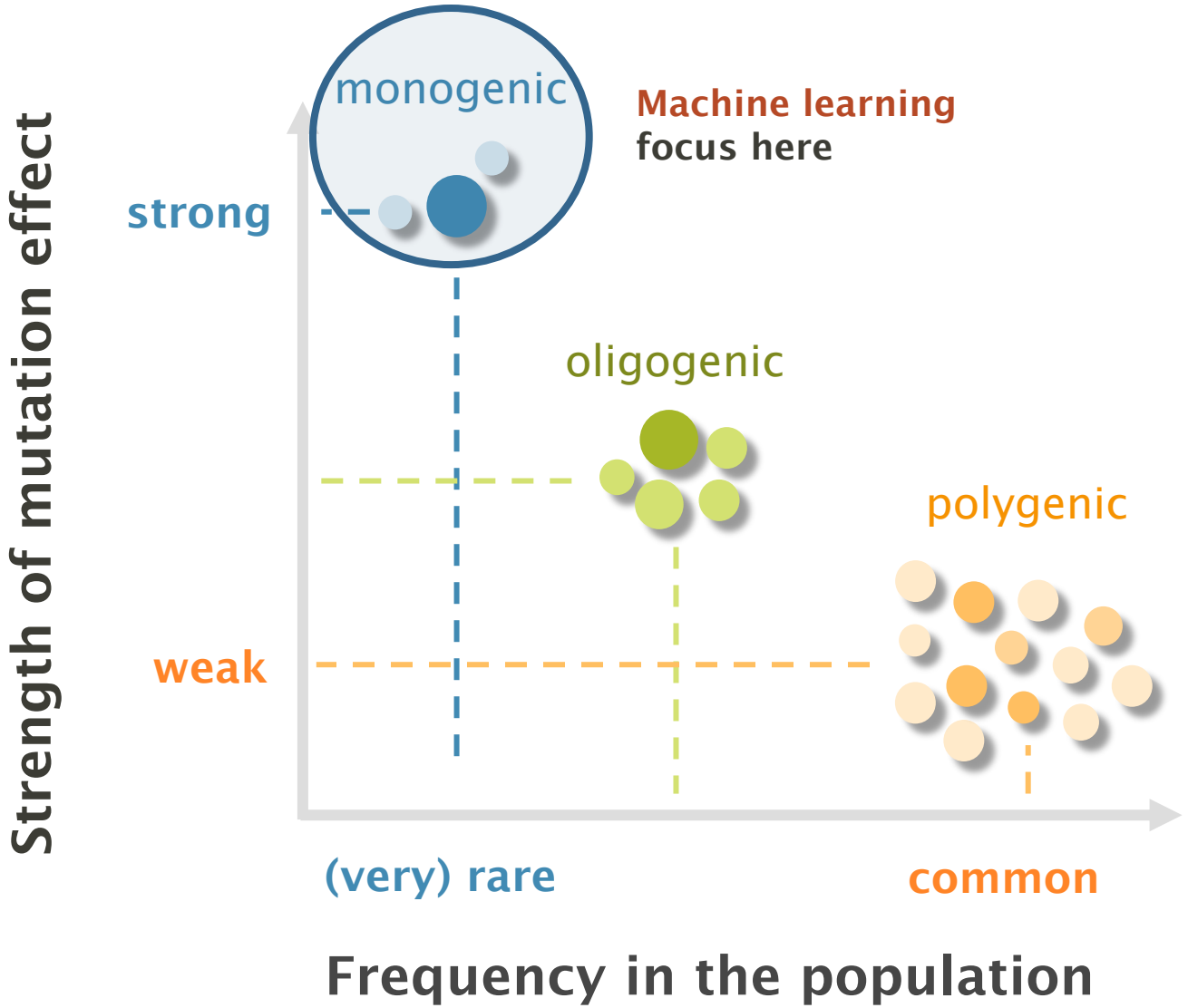
INTERPRETABILITY IS IMPORTANT



An interpretable model can be more fair, reliable and trustworthy.

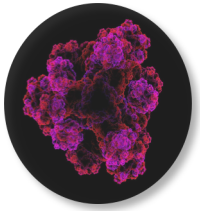
Source: https://www.infoq.com/presentations/machine-learning-gdpr/

MACHINE LEARNING FOCUS ON MONOGENIC



Source: McCarthy, *et al.* (2008). *Nature reviews genetics*. 9(5), 356.

MACHINE LEARNING IN VARIANT PREDICTION



"Normal" …AGTGCGGTG… …SAV…

"Disease" …AGTGAGGTG… …SEV…

Mutation Taster

PolyPhen-2

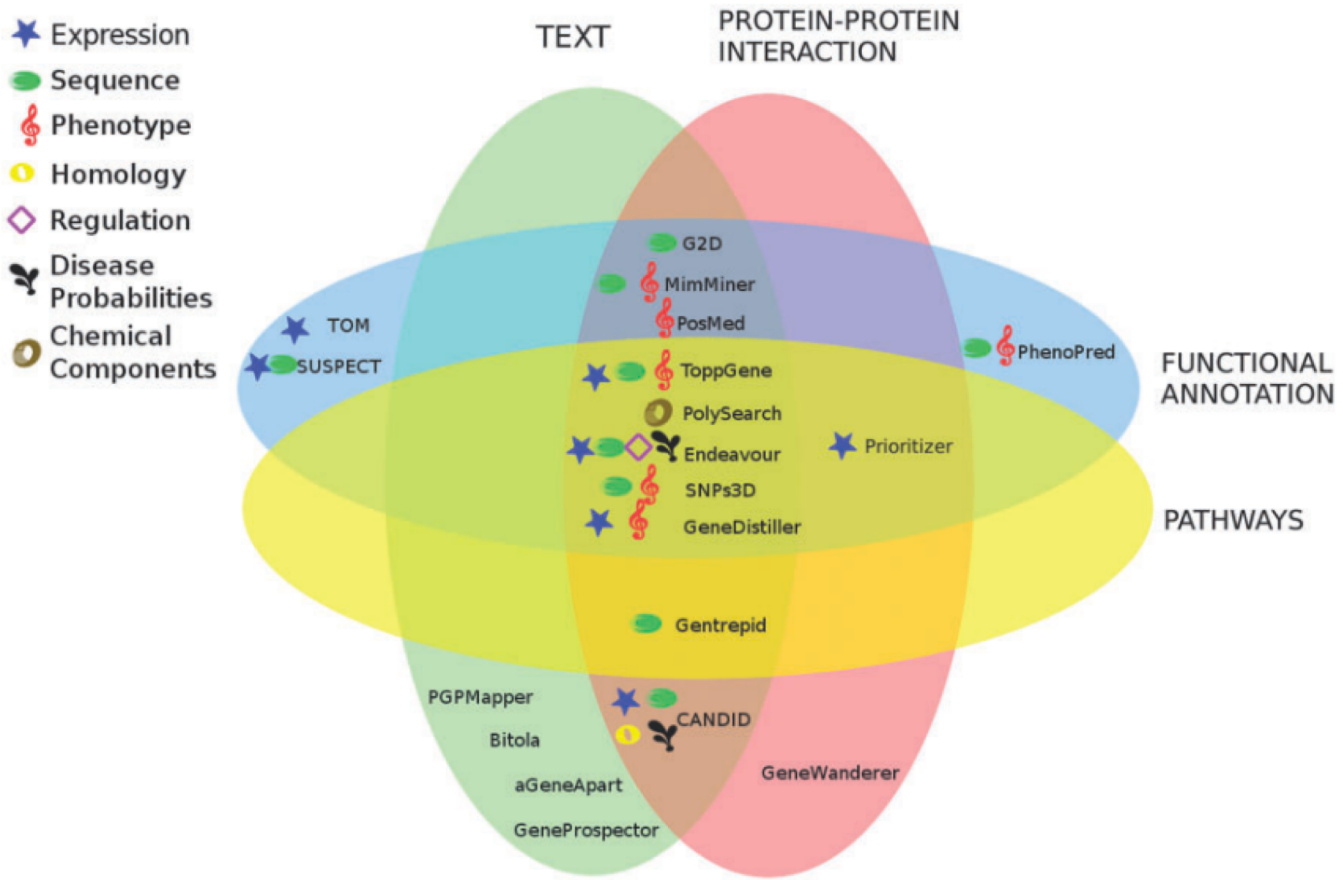
CADD

eXtasy Variant Prioritization

SIFT

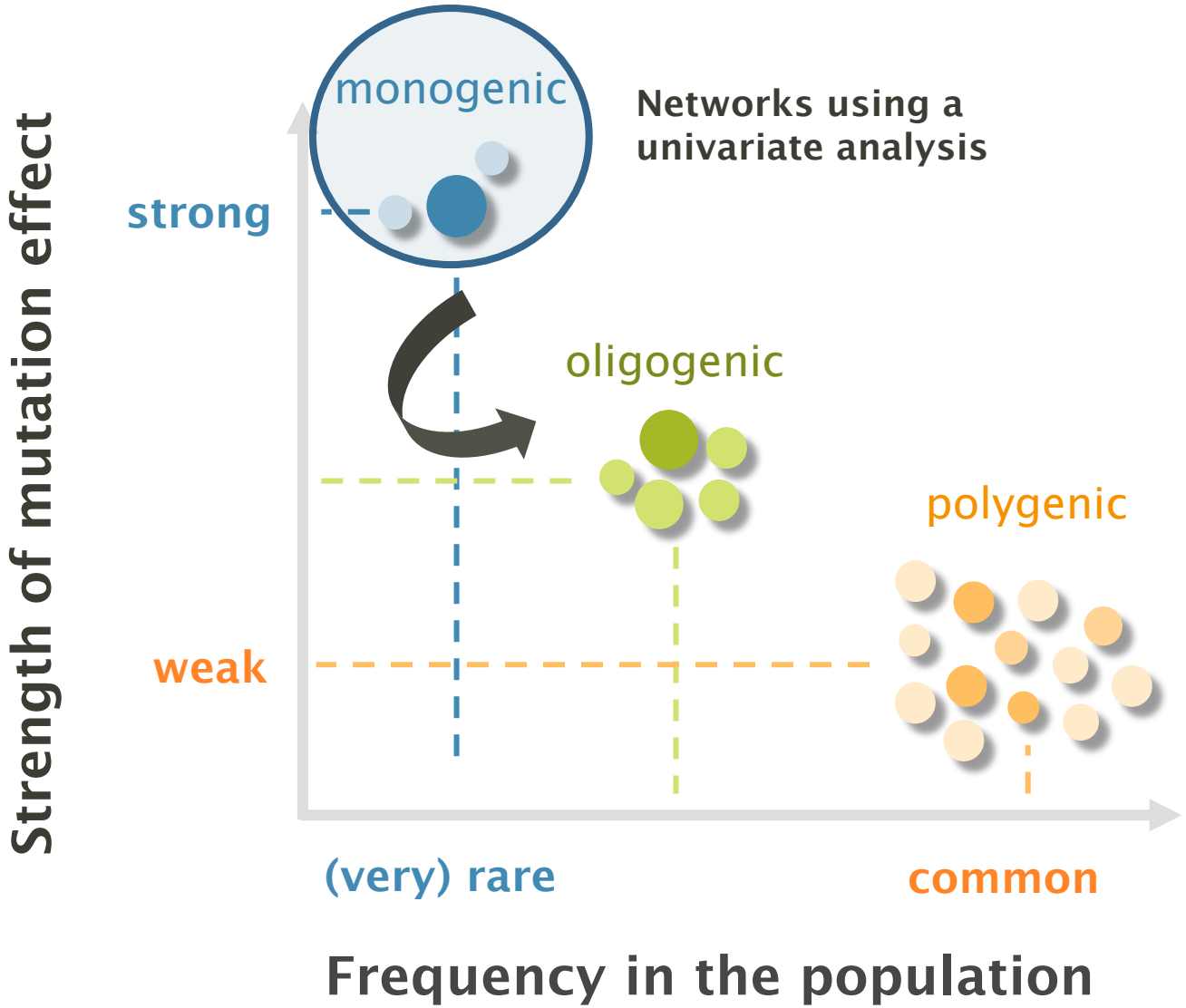
DEOGEN

MACHINE LEARNING IN GENE PREDICTION



Source: Tranchevent, *et al.* (2011). *Briefings in Bioinformatics*. **12**(1), 22.

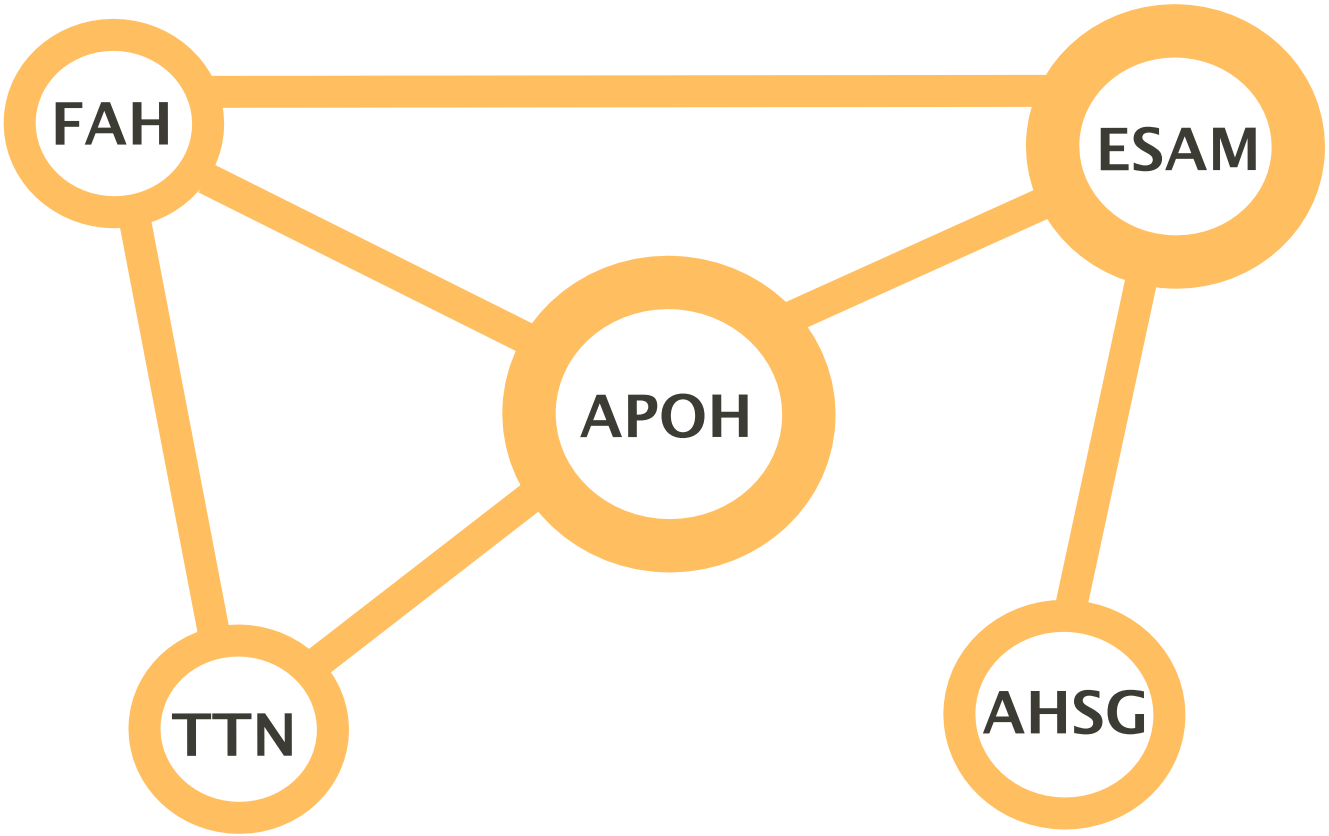
FROM MONOGENIC TO OLIGOGENIC



Source: McCarthy, *et al.* (2008). *Nature reviews genetics*. 9(5), 356.

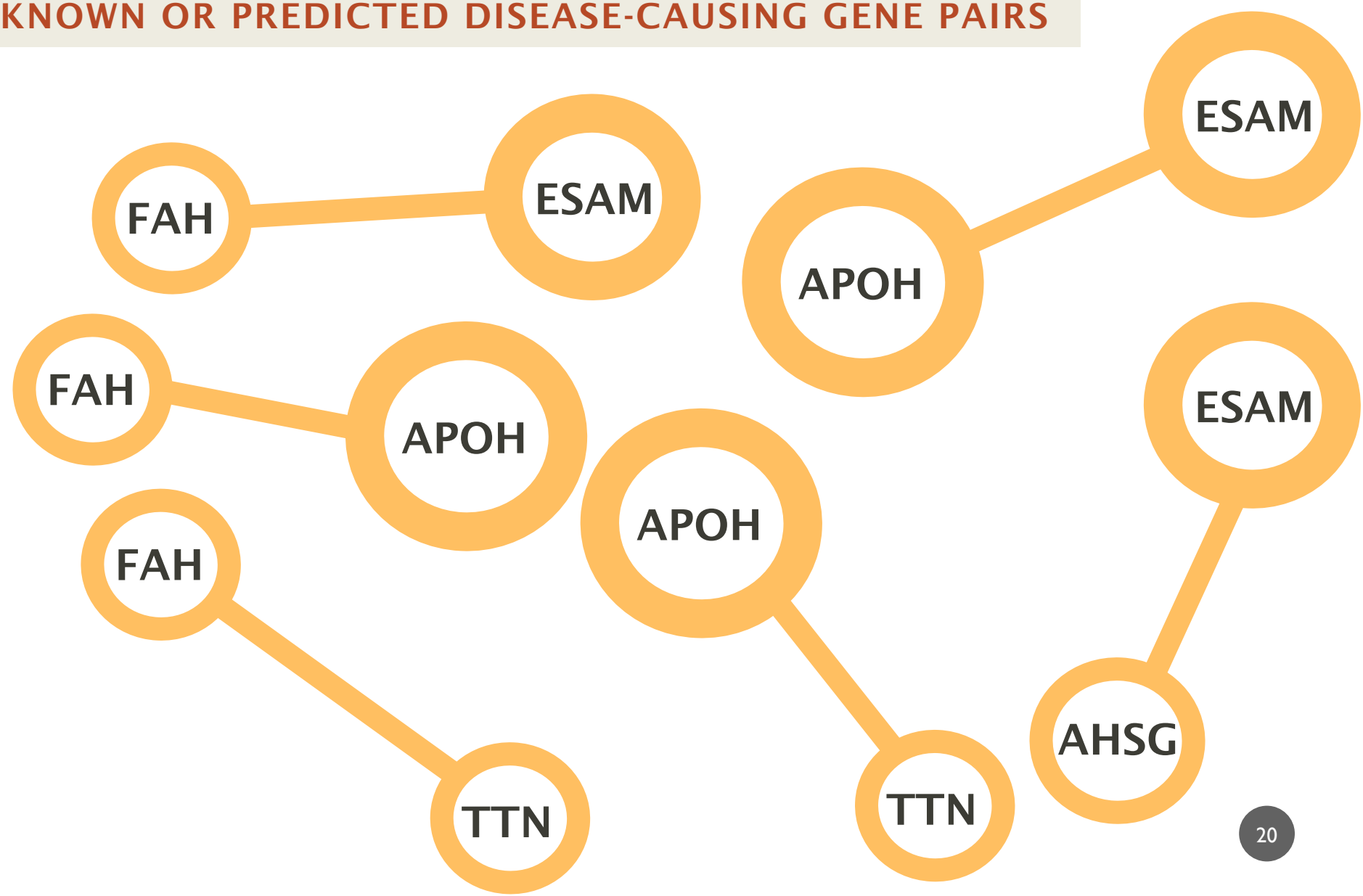
NETWORKS: THE NODE TO EDGE APPROACH

BASED ON PPIs, CO-EXPRESSION, PRESENCE OF SEVERE SNPs



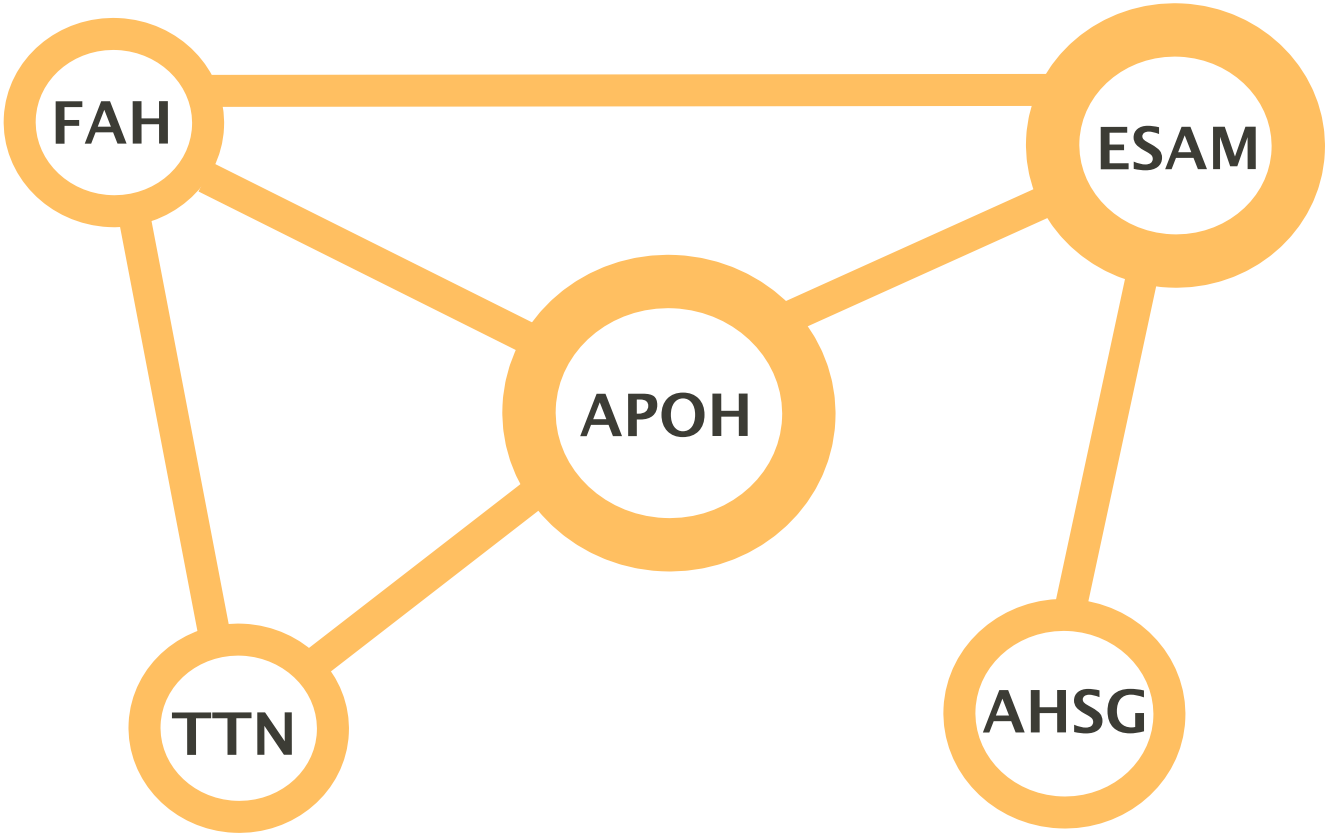
DIFFERENT APPROACH: FROM EDGES TO NODES

KNOWN OR PREDICTED DISEASE-CAUSING GENE PAIRS



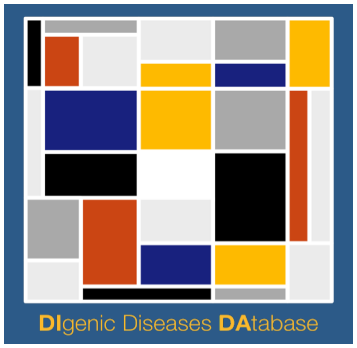
DIFFERENT APPROACH: FROM EDGES TO NODES

BASED ON KNOWN OR PREDICTED KNOWLEDGE

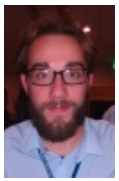


HYPOTHESIS 1:
Sufficient cases **exist**, where mutations in **two genes** explain **better** the phenotype of a patient than a mutation in **one gene alone**.

DIDA: THE DIGENIC DISEASES DATABASE



http://dida.ibsquare.be


Andrea Gazzo


Dorien Daneels


Claudio Reggiani

bi-locus variant combination

BBS2

A T C A A G T A C C G

C

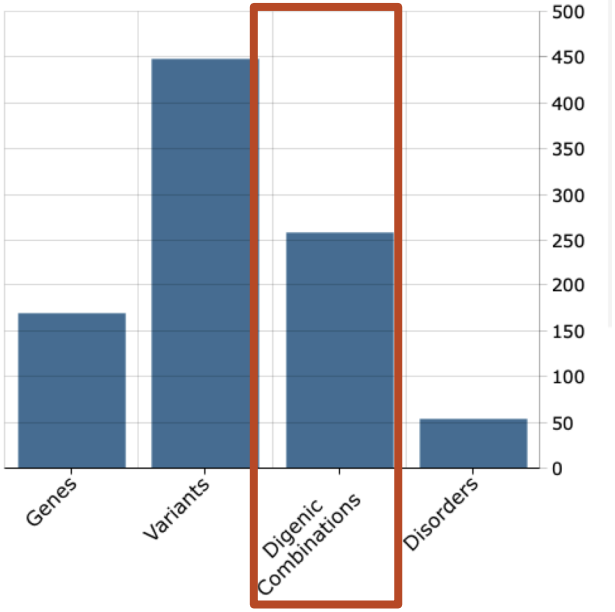
MKKS

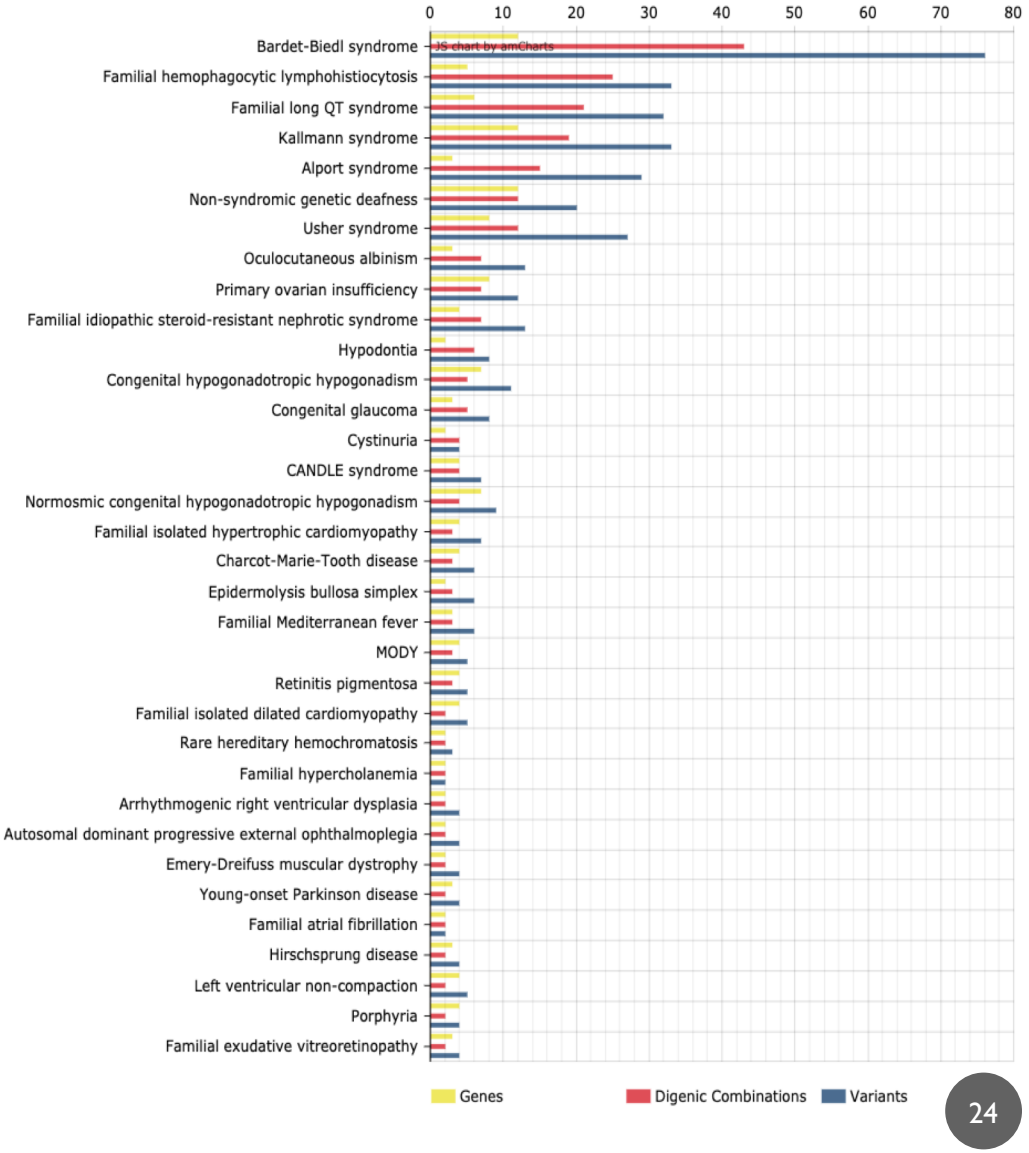
T G A A C C C T G G G

A

Bardet-Biedl syndrome

258 COMBINATIONS, 55 DIGENIC DISEASES





Source: http://dida.ibsquare.be

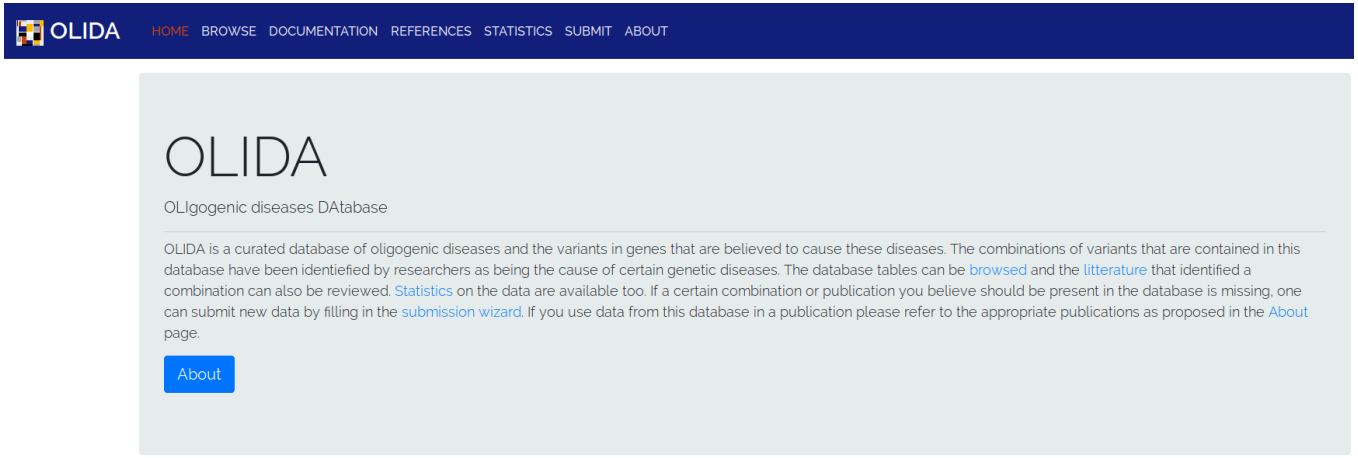
THE COMBINATIONS PAGE IN DIDA

A bi-locus variant combination

	Gene A				Gene B								
ID	Name	Allele 1 protein change	Allele 2 protein change	Zygosity	Name	Allele 1 protein change	Allele 2 protein change	Zygosity	Disease name (ORPHANET)	Oligogenic effect	Familial evidence	Functional evidence	Gene relationship
dd001	KCNQ1	p.(A341E)	wild type	Heterozygote	KCNH2	N/A	wild type	Heterozygote	Familial long QT syndrome	CO	YES	NO	indirectly interacting, pathway membership, similar function, co-expression
dd002	GJB3	p.(N166S)	wild type	Heterozygote	GJB2	p.(L79Cfs*3)	wild type	Heterozygote	Non-syndromic genetic deafness	TD	YES	NO	indirectly interacting, pathway membership, similar function
dd003	GJB3	p.(A194T)	wild type	Heterozygote	GJB2	p.(L79Cfs*3)	wild type	Heterozygote	Non-syndromic genetic deafness	TD	YES	NO	indirectly interacting, pathway membership, similar function
dd004	GJB3	p.(A194T)	wild type	Heterozygote	GJB2	p.(H100Rfs*14)	wild type	Heterozygote	Non-syndromic genetic deafness	TD	YES	NO	indirectly interacting, pathway membership, similar function
dd005	FOXI1	p.(G258E)	wild type	Heterozygote	SLC26A4	p.(E29Q)	wild type	Heterozygote	Non-syndromic genetic deafness	TD	YES	YES	co-expression

Source: http://dida.ibsquare.be

OLIDA IS COMING SOON



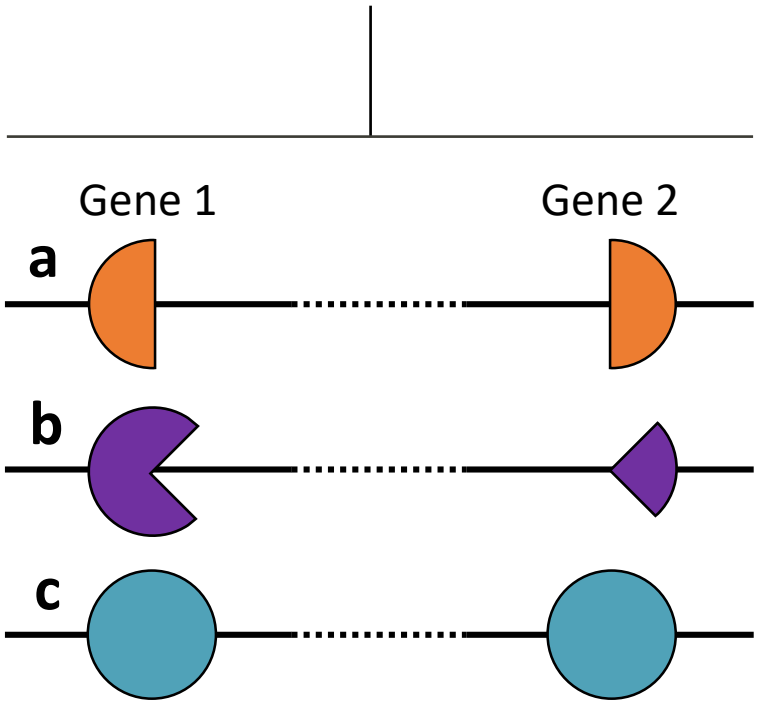
Arnau Dillen

Charlotte Nachtegael

- Adapted for oligogenic cases
- Adapted for CNVs
- Improved paper curation
- Faster data submission

MAIN TYPES OF BI-LOCUS COMBINATIONS

A bi-locus model explains better the phenotype



True Digenic

Monogenic + Modifier

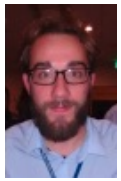
Dual Diagnosis

Problem: 1/3 of the data is not classified with a bi-locus effect

Source: Versbraegen, *et al.* (2019). *Artificial Intelligence in Medicine.* **99**, 101690

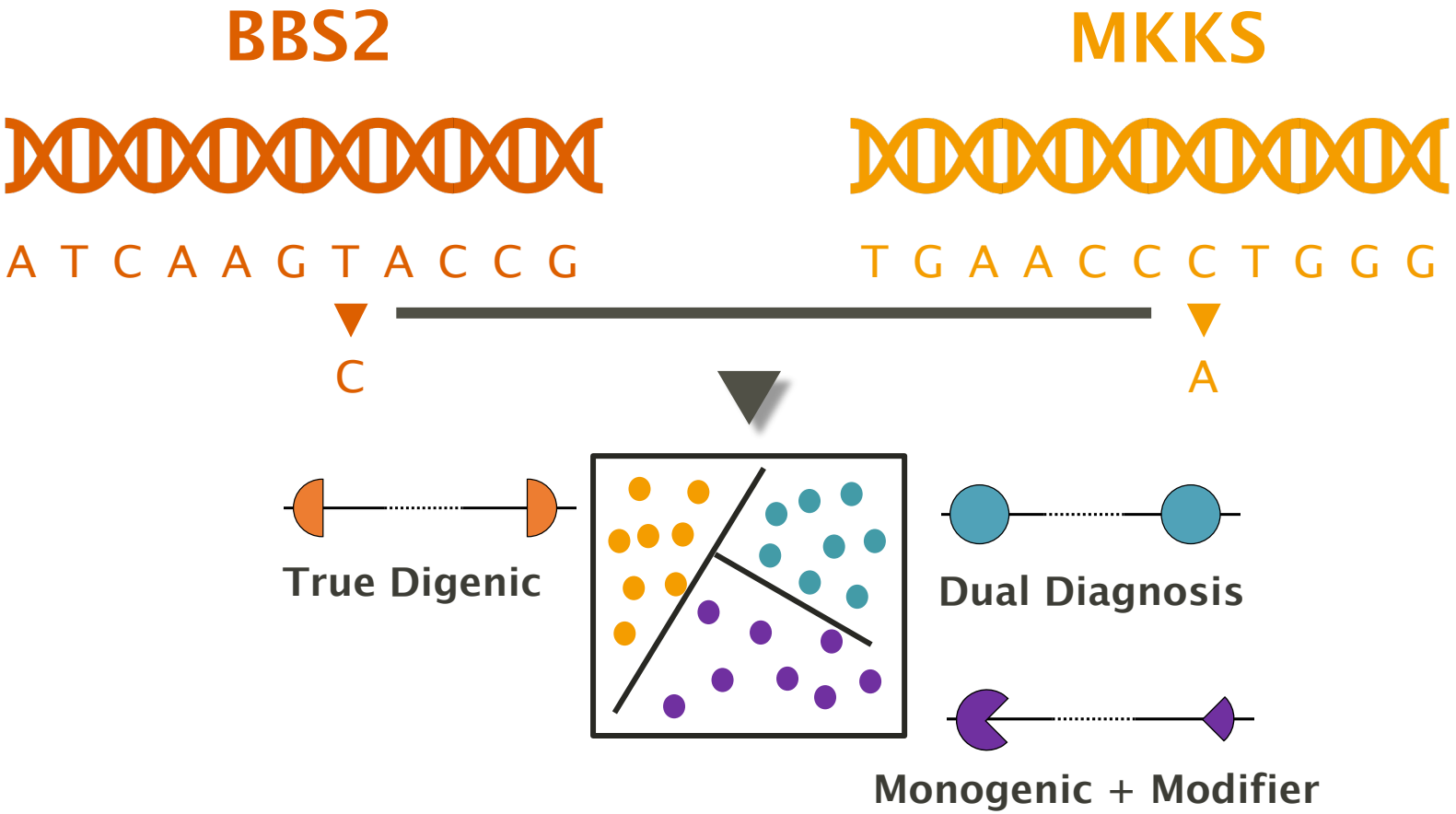
HYPOTHESIS 2:
It is possible to differentiate between different types of bi-locus combinations using machine learning.

THE BI-LOCUS EFFECT PREDICTOR


Andrea Gazzo


Aziz Fouché

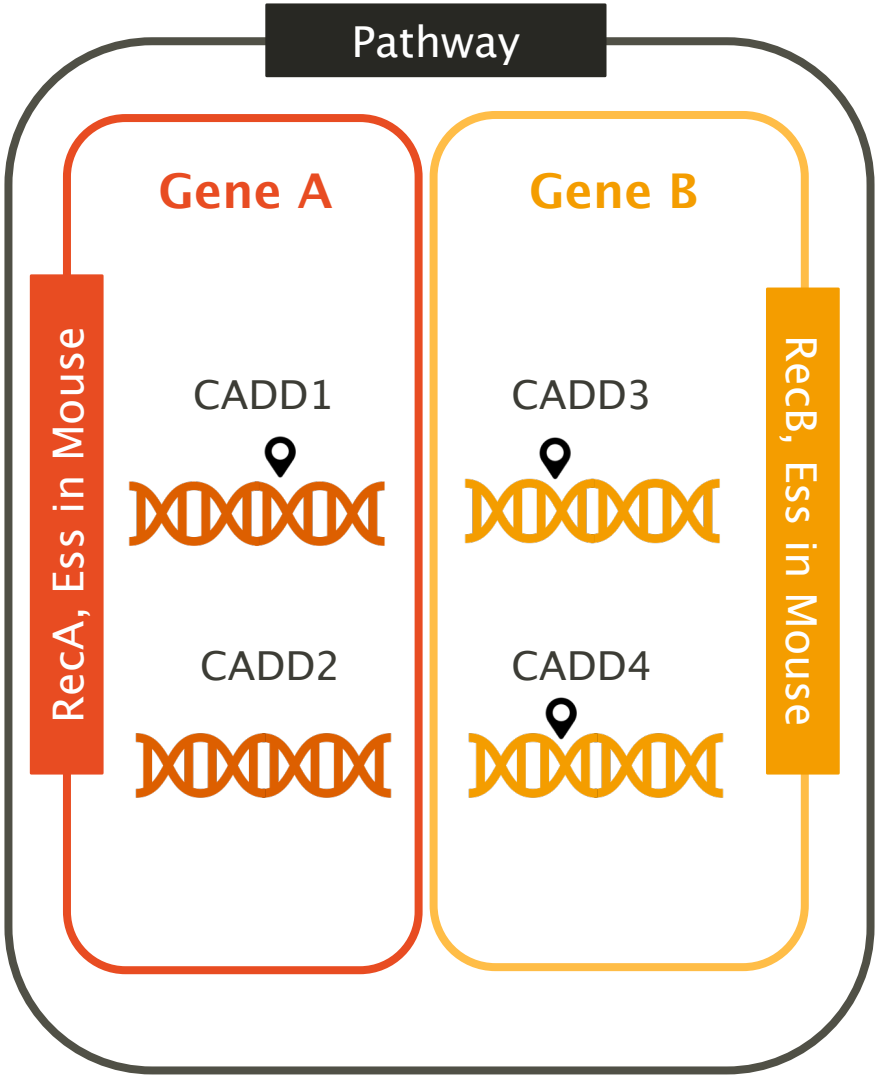

Nassim Versbraegen



Source: Versbraegen, *et al.* (2019). *Artificial Intelligence in Medicine.* **99**, 101690

THE BI-LOCUS EFFECT PREDICTOR

USING A MULTI-LAYER ANNOTATION



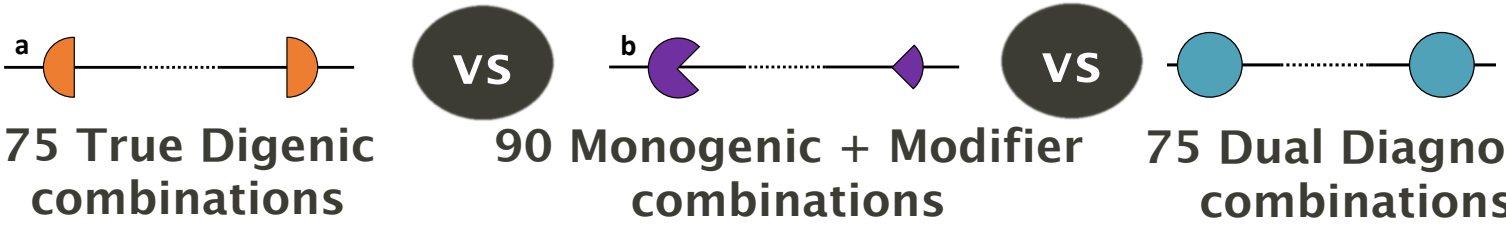
[0.52, -10, …. , 0.82, 1]

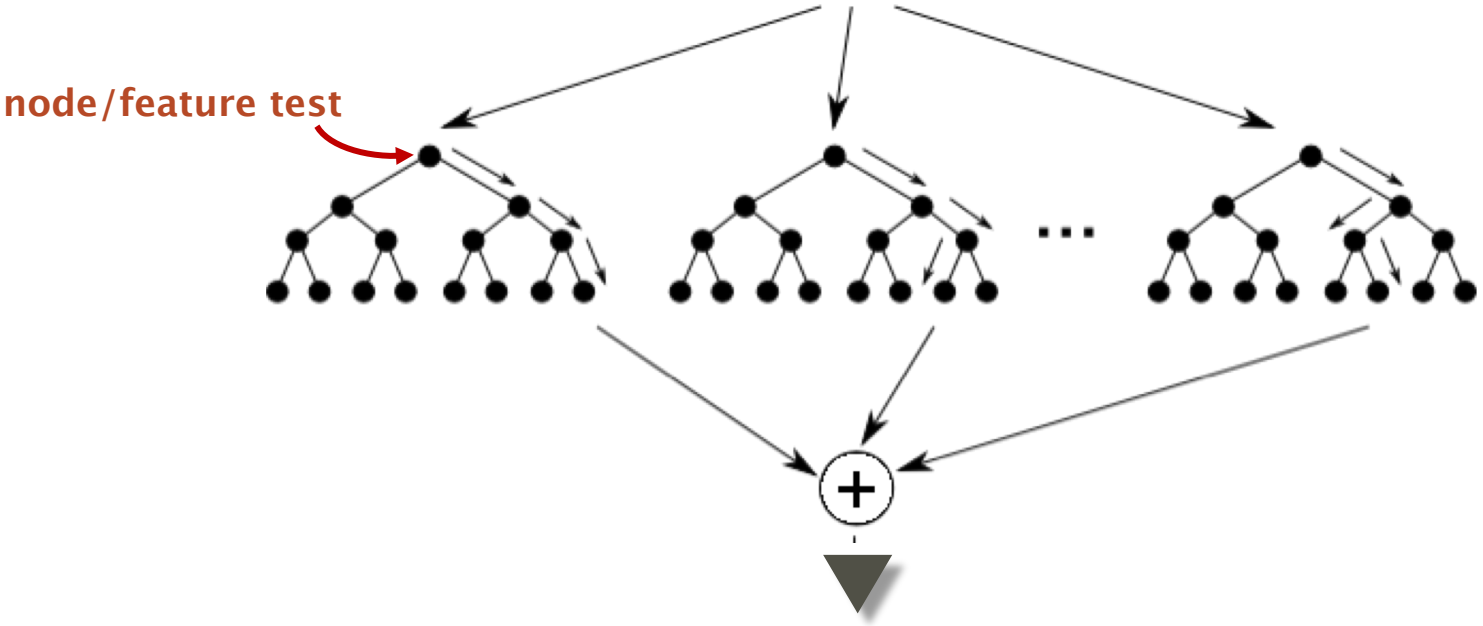
Vector of 9 elements (features)

Source: Versbraegen, *et al.* (2019). *Artificial Intelligence in Medicine.* **99**, 101690

THE BI-LOCUS EFFECT PREDICTOR

A RANDOM FOREST MODEL

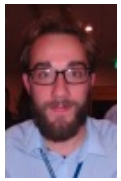




Source: Versbraegen, *et al.* (2019). *Artificial Intelligence in Medicine.* **99**, 101690

THE BI-LOCUS EFFECT PREDICTOR

A MODEL WITH GOOD PERFORMANCE



It's easier to differentiate dual molecular diagnosis from the other two classes.

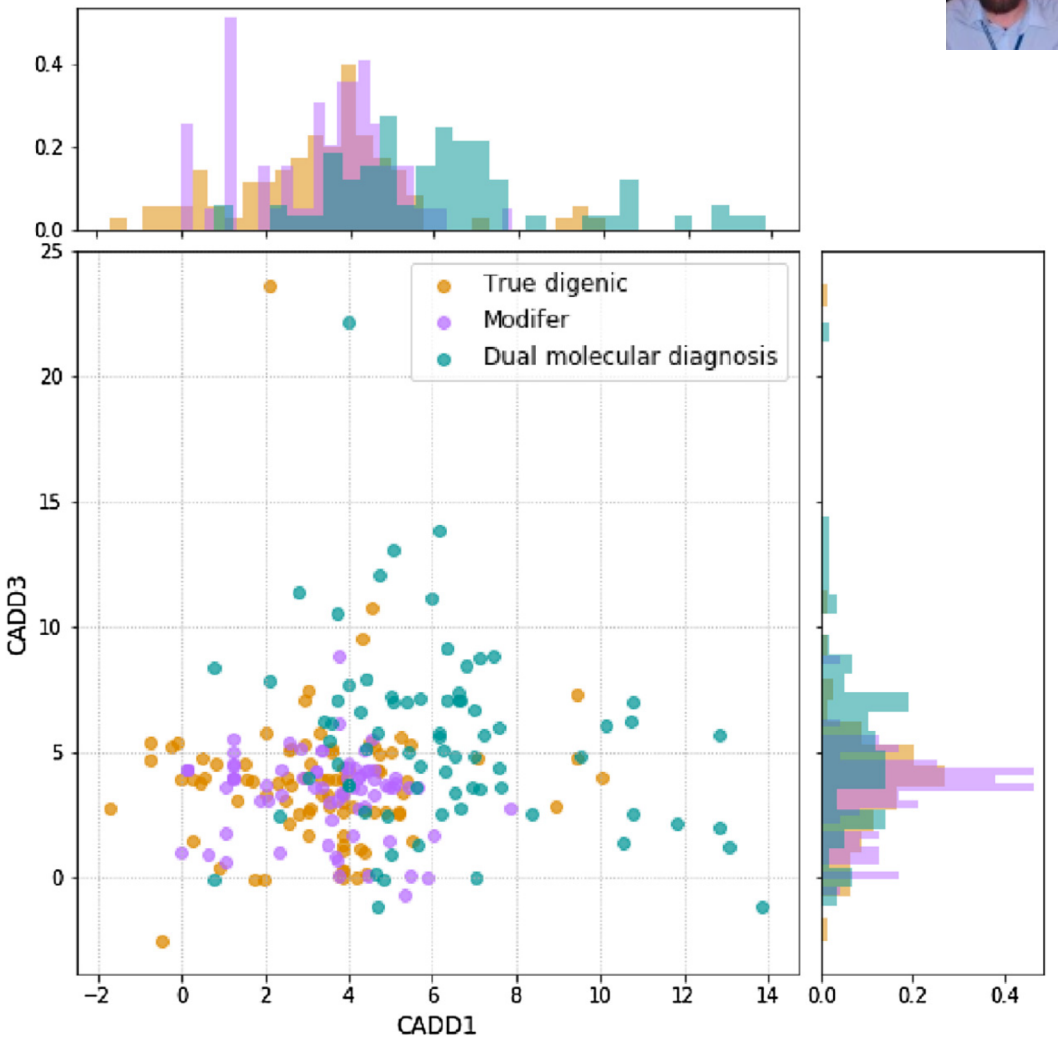
Class	Sensitivity	Specificity
Dual Molecular Diagnosis	0.8	0.79
Monogenic + Modifier	0.57	0.65
True Digenic	0.7	0.65

Source: Versbraegen, *et al.* (2019). *Artificial Intelligence in Medicine.* **99**, 101690

THE BI-LOCUS EFFECT PREDICTOR

DMDs DIFFERENTIATED (CADD, PATHWAY)

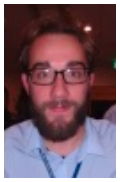




Source: Versbraegen, *et al.* (2019). *Artificial Intelligence in Medicine.* **99**, 101690

THE BI-LOCUS EFFECT PREDICTOR

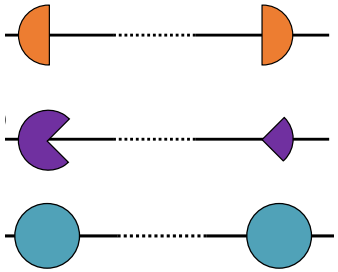
INTERPRETABILITY WITH GAME THEORY



Pay-off cooperative game using the set F of 9 features as agents

Reward: sensitivity, specificity

Testing all possible $2^9 - 1$ feature coalitions S belonging to F



x

sensitivity

specificity

=

6 games,
6 rewards

▶ **reward v(S)**: the geometric mean

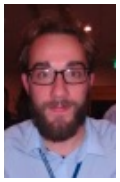
▼

Shapley value ϕ_F:
a relative value on how much an agent contributes to a **v(S)**

Source: Versbraegen, *et al.* (2019). *Artificial Intelligence in Medicine.* **99**, 101690

THE BI-LOCUS EFFECT PREDICTOR

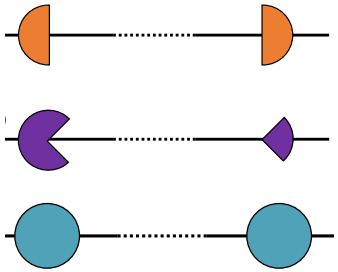
INTERPRETABILITY WITH GAME THEORY



Pay-off cooperative game using **all pairs of F features** as agents

Reward: **sensitivity**, **specificity**

Testing all possible feature coalitions S of pairs F_i and F_j



x

sensitivity

specificity

= 6 games, 6 rewards ▶ **reward v(S)**: the geometric mean

▼

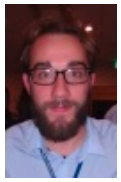
joint contribution $\gamma_{i,j}$

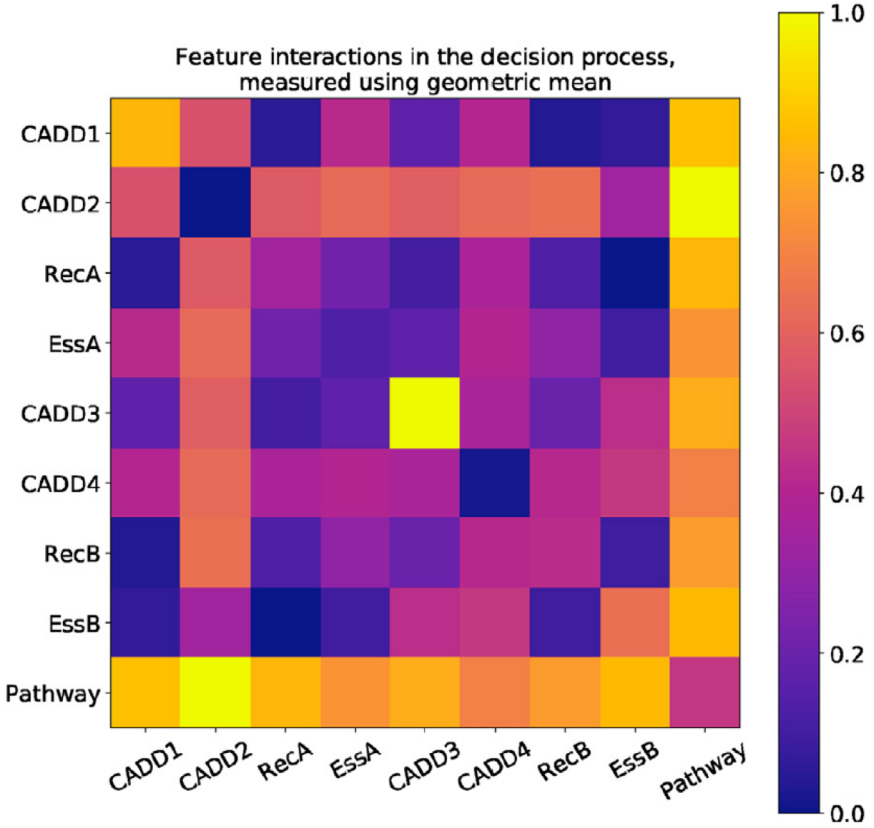
A feature pair can be redundant, complementary or synergistic

Source: Versbraegen, *et al.* (2019). *Artificial Intelligence in Medicine.* **99**, 101690

THE BI-LOCUS EFFECT PREDICTOR

INTERPRETABILITY WITH GAME THEORY





- High synergy of pathway with almost any feature
- Synergy of CADD2 and CADD4 with pathway and gene recessiveness / essentiality

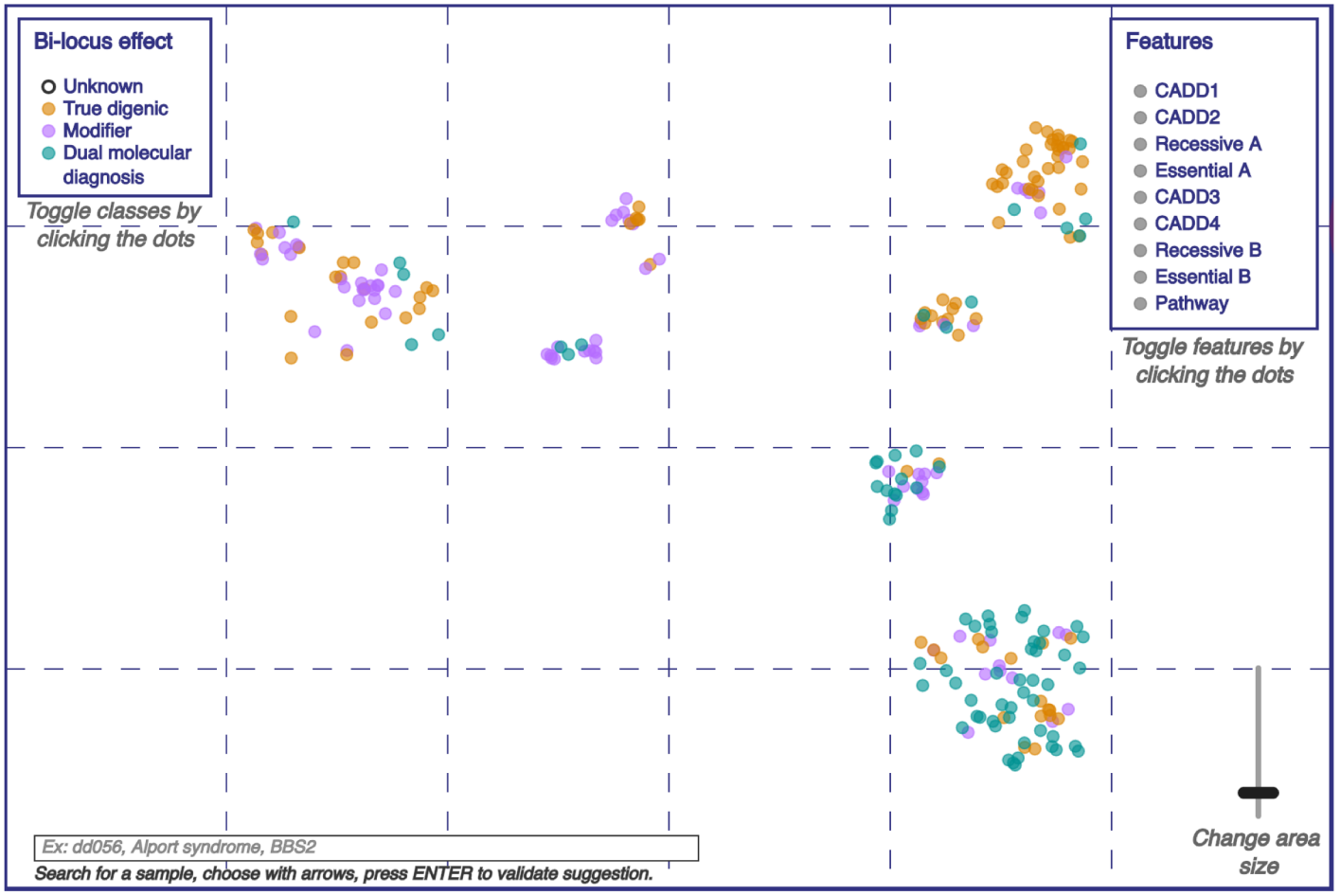
Source: Versbraegen, *et al.* (2019). *Artificial Intelligence in Medicine.* **99**, 101690

THE BI-LOCUS EFFECT PREDICTOR

CLUSTERING DIDA BI-LOCUS COMBINATIONS



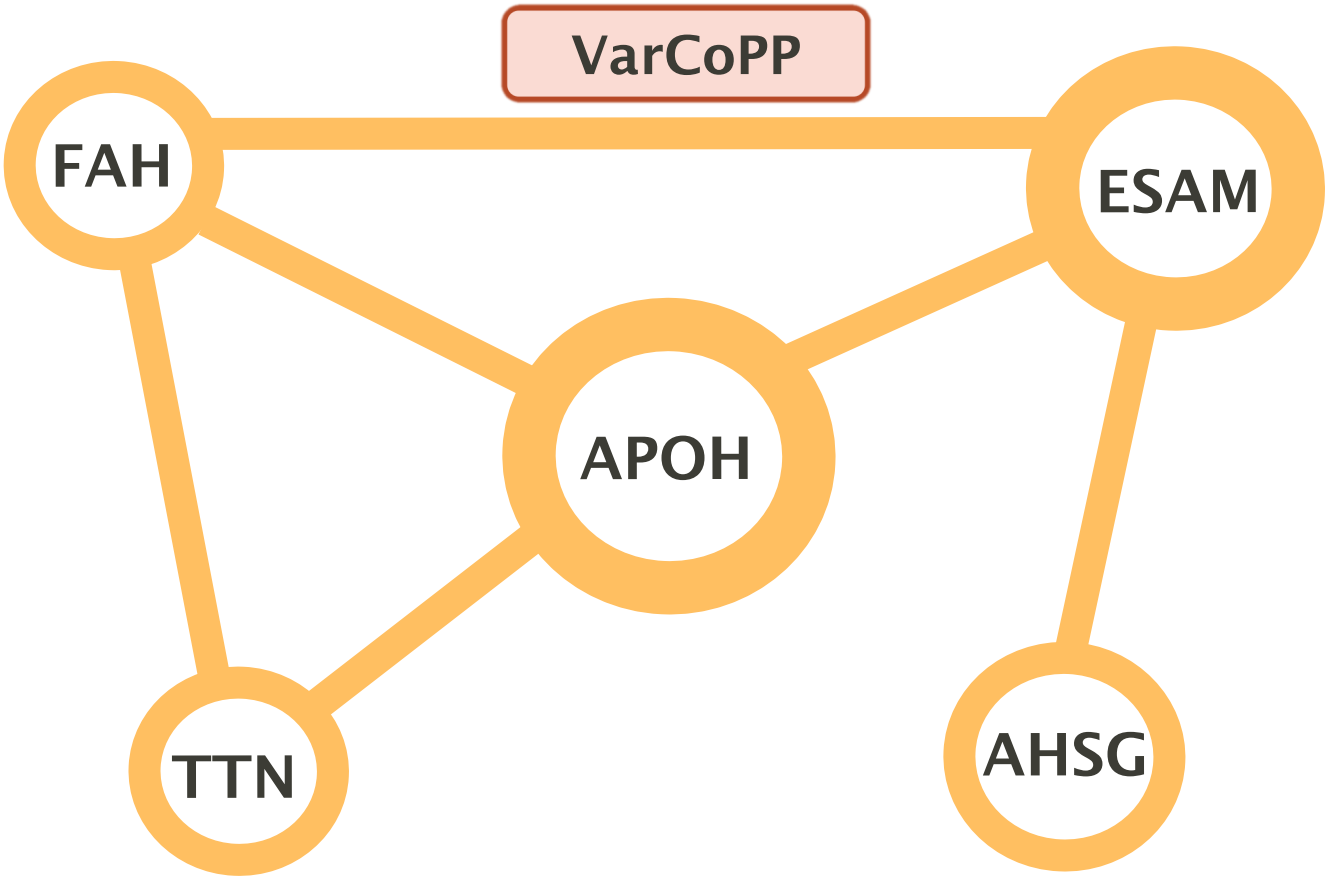
http://bespace.ibsquare.be/



Source: Versbraegen, *et al.* (2019). *Artificial Intelligence in Medicine.* **99**, 101690

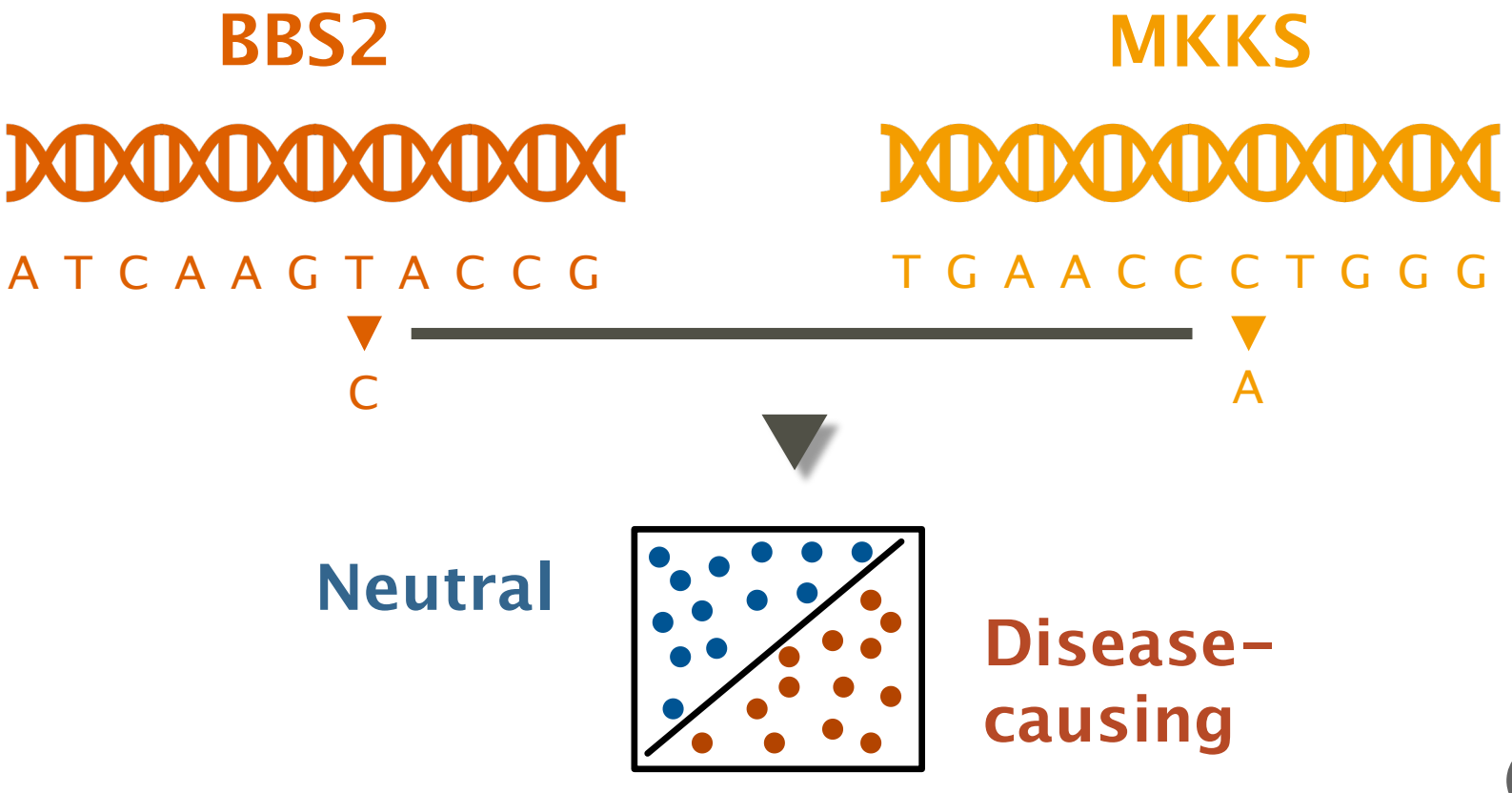
HYPOTHESIS 3:
It is possible to differentiate between disease-causing and neutral combinations using machine learning.

DIFFERENT APPROACH: FROM EDGES TO NODES



VARCOPP: THE PATHOGENICITY PREDICTOR

Sofia Papadimitriou



Source: Versbraegen, *et al.* (2019). *Artificial Intelligence in Medicine.* **99**, 101690

VARCOPP: THE PATHOGENICITY PREDICTOR

THE DATA

DIgenic Diseases DAtabase

vs

1000 Genomes
A Deep Catalog of Human Genetic Variation

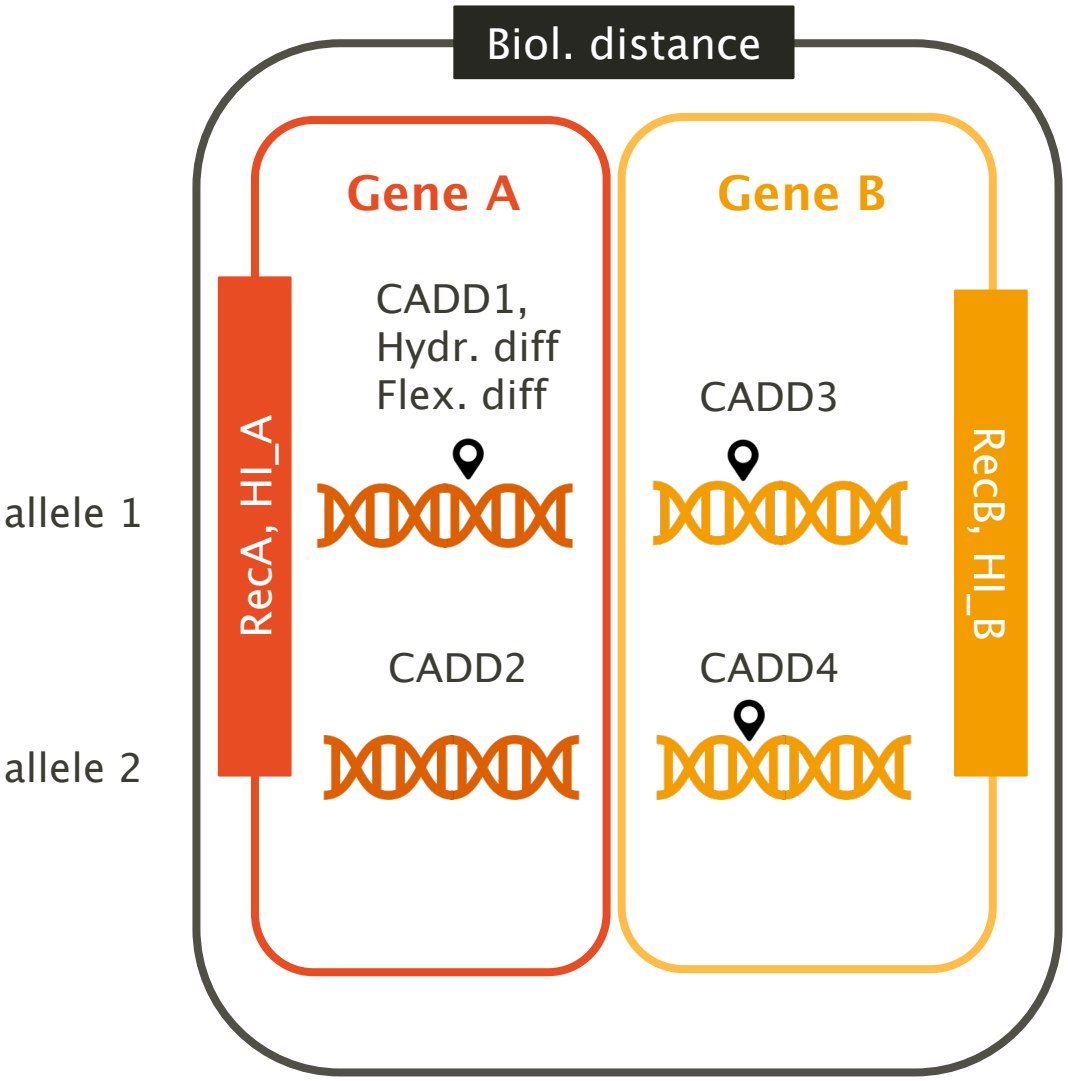
213 bi-locus combinations

2500 individuals trillions of combinations

Source: Papadimitriou, *et al.* (2019). *PNAS.* **116**(24), 11878

VARCOPP: THE PATHOGENICITY PREDICTOR

MULTILAYER ANNOTATION



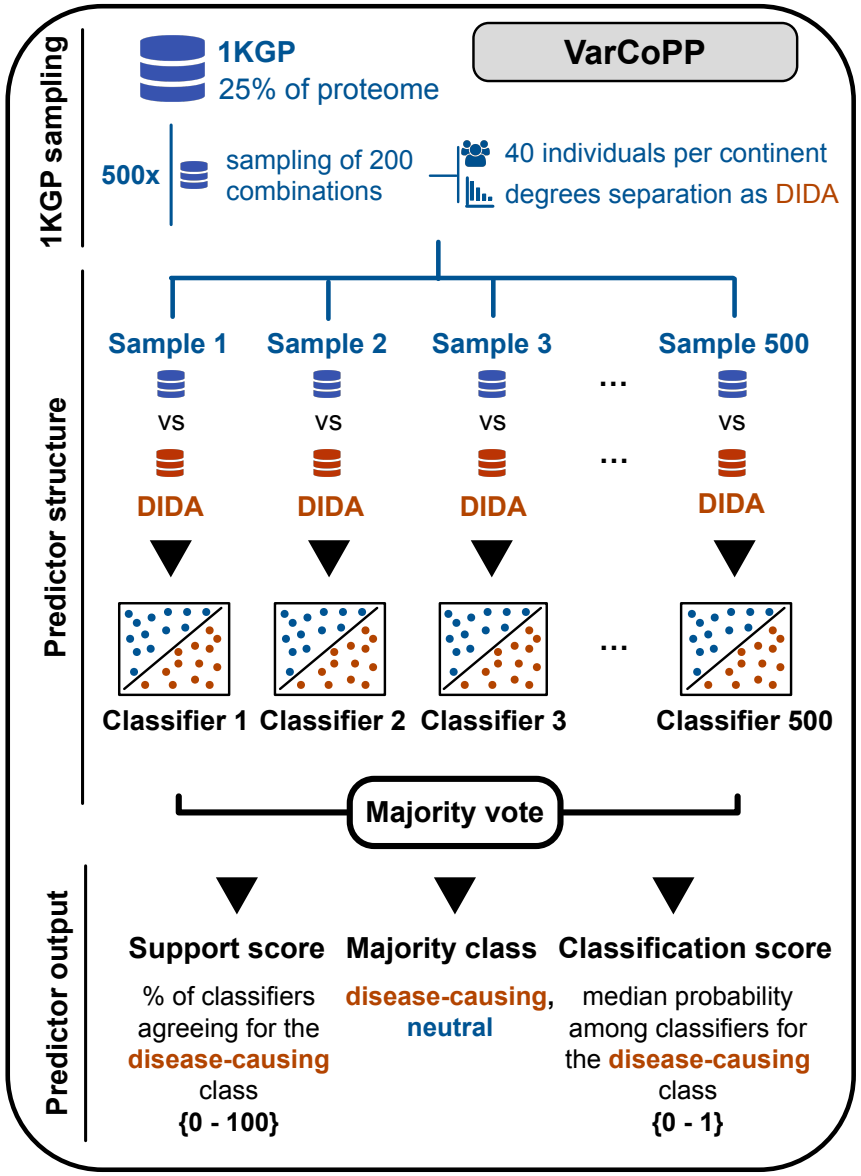
[0.52, -10, …. , 0.82, 1]

Vector of 11 elements (features)

Source: Papadimitriou, *et al.* (2019). *PNAS.* **116**(24), 11878

VARCOPP: THE PATHOGENICITY PREDICTOR

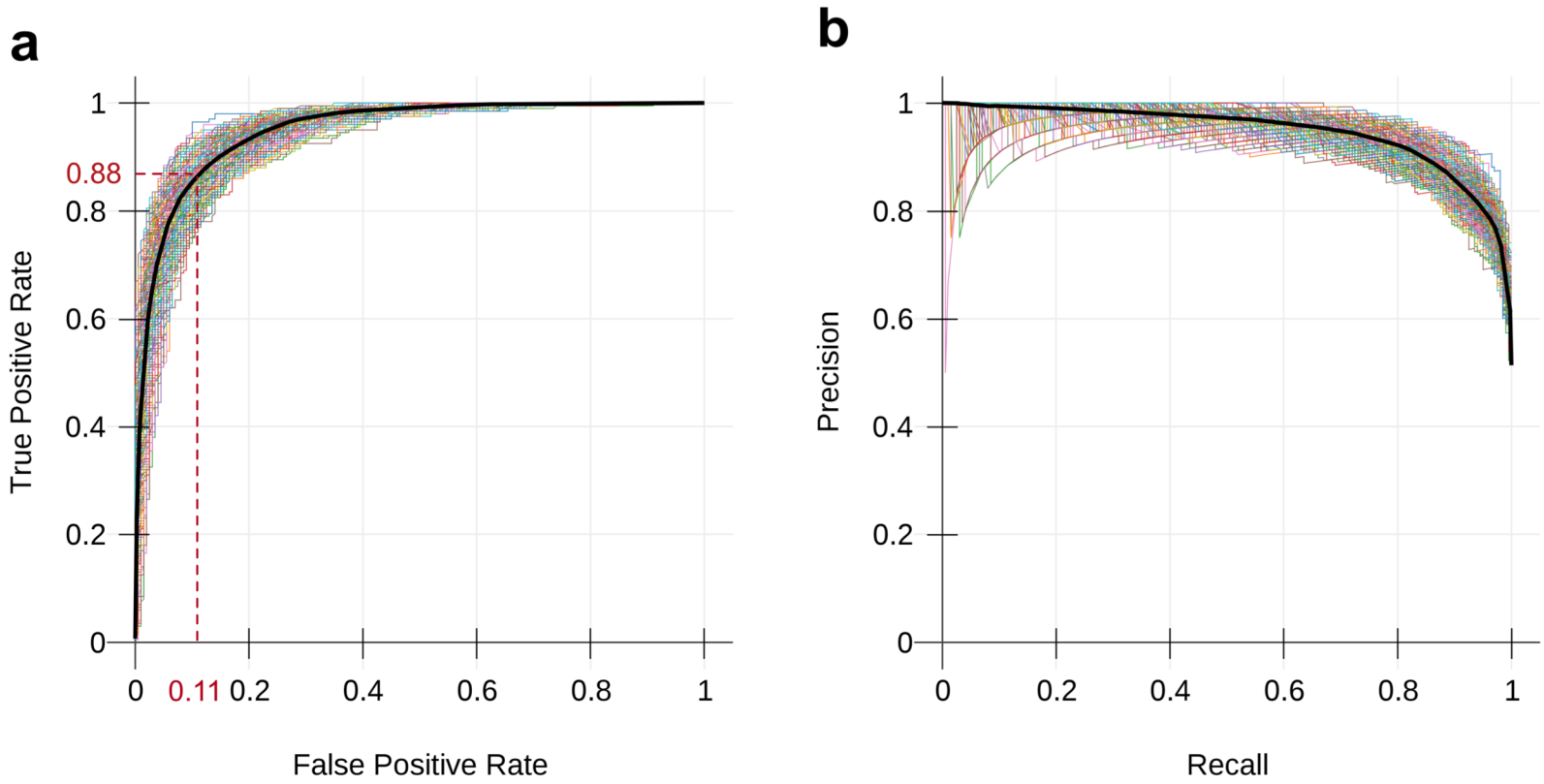
THE MODEL



Source: Papadimitriou, *et al.* (2019). *PNAS.* **116**(24), 11878

VARCOPP: THE PATHOGENICITY PREDICTOR

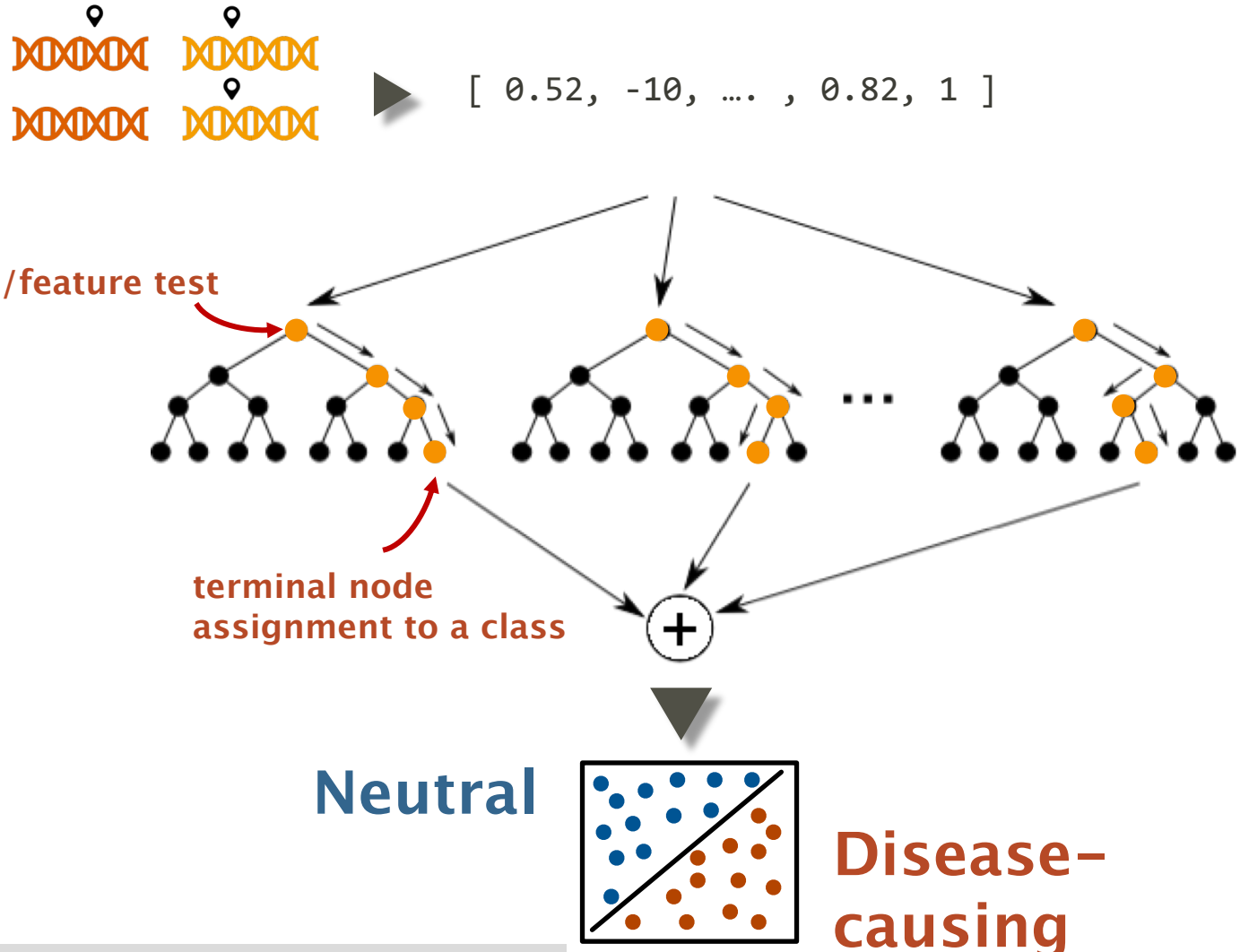
0.88 ACCURACY, 0.74 MCC



Source: Papadimitriou, *et al.* (2019). *PNAS.* **116**(24), 11878

THE BI-LOCUS EFFECT PREDICTOR

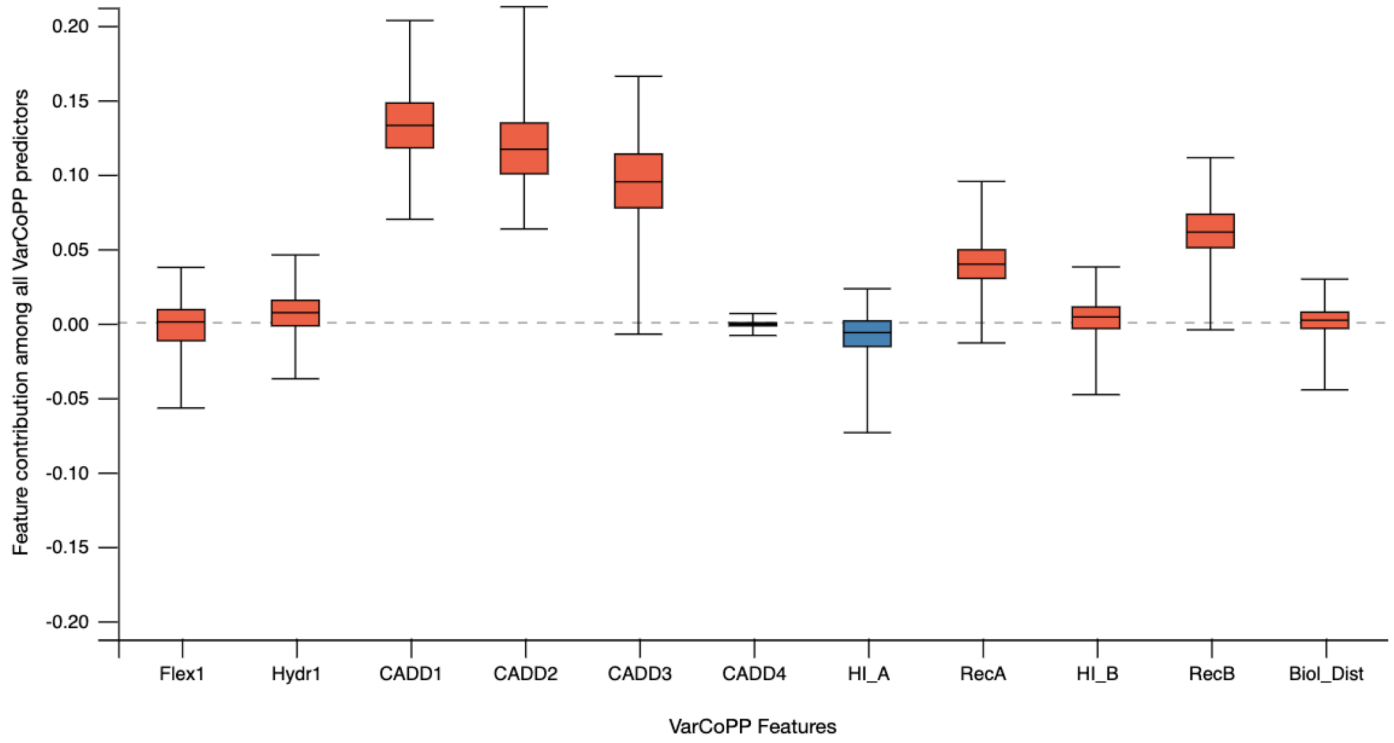
INTERPRETABILITY WITH RANDOM FOREST



Source: Papadimitriou, *et al.* (2019). *PNAS.* **116**(24), 11878

VARCOPP: THE PATHOGENICITY PREDICTOR

INTERPRETABLE PREDICTIONS



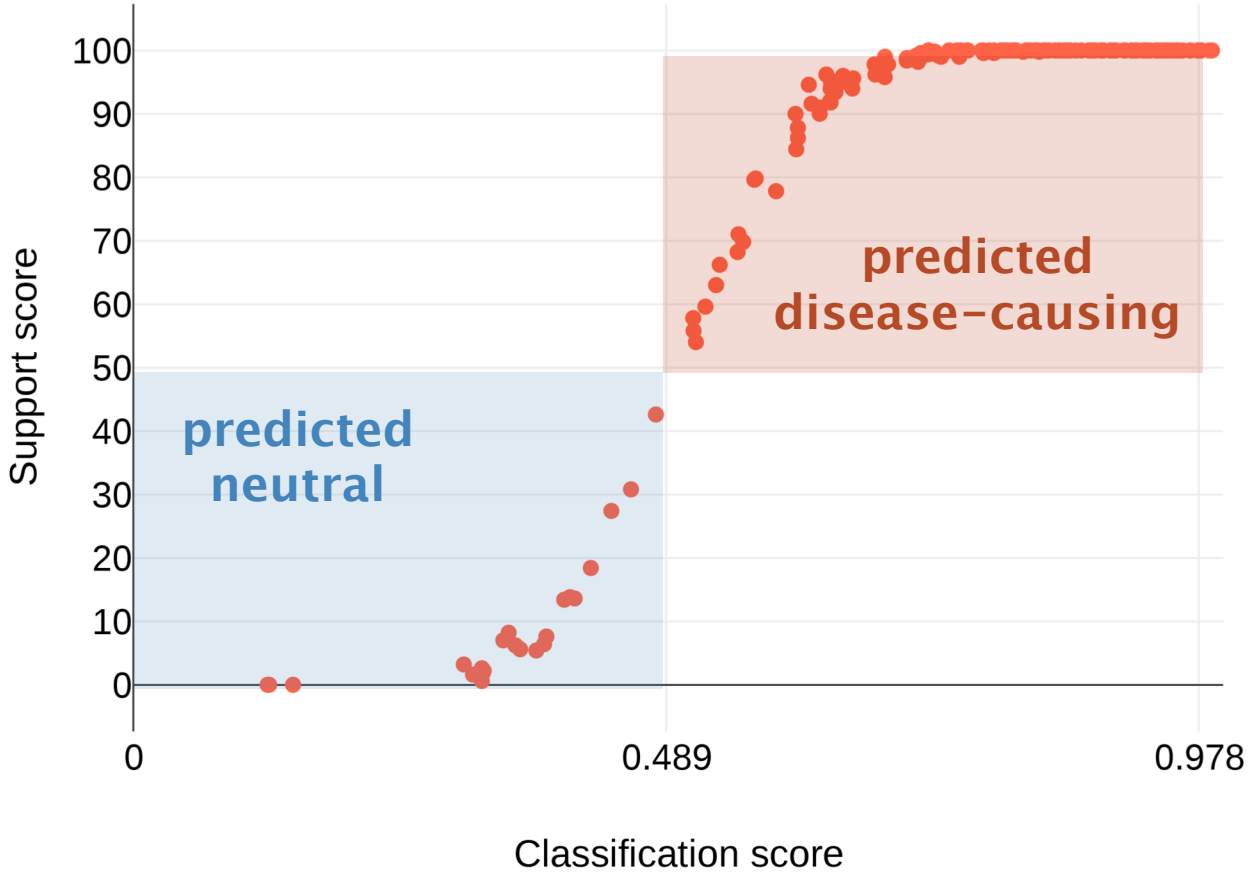
Preference for the disease-causing class

Preference for the neutral class

Source: Papadimitriou, *et al.* (2019). *PNAS.* **116**(24), 11878

VARCOPP: THE PATHOGENICITY PREDICTOR

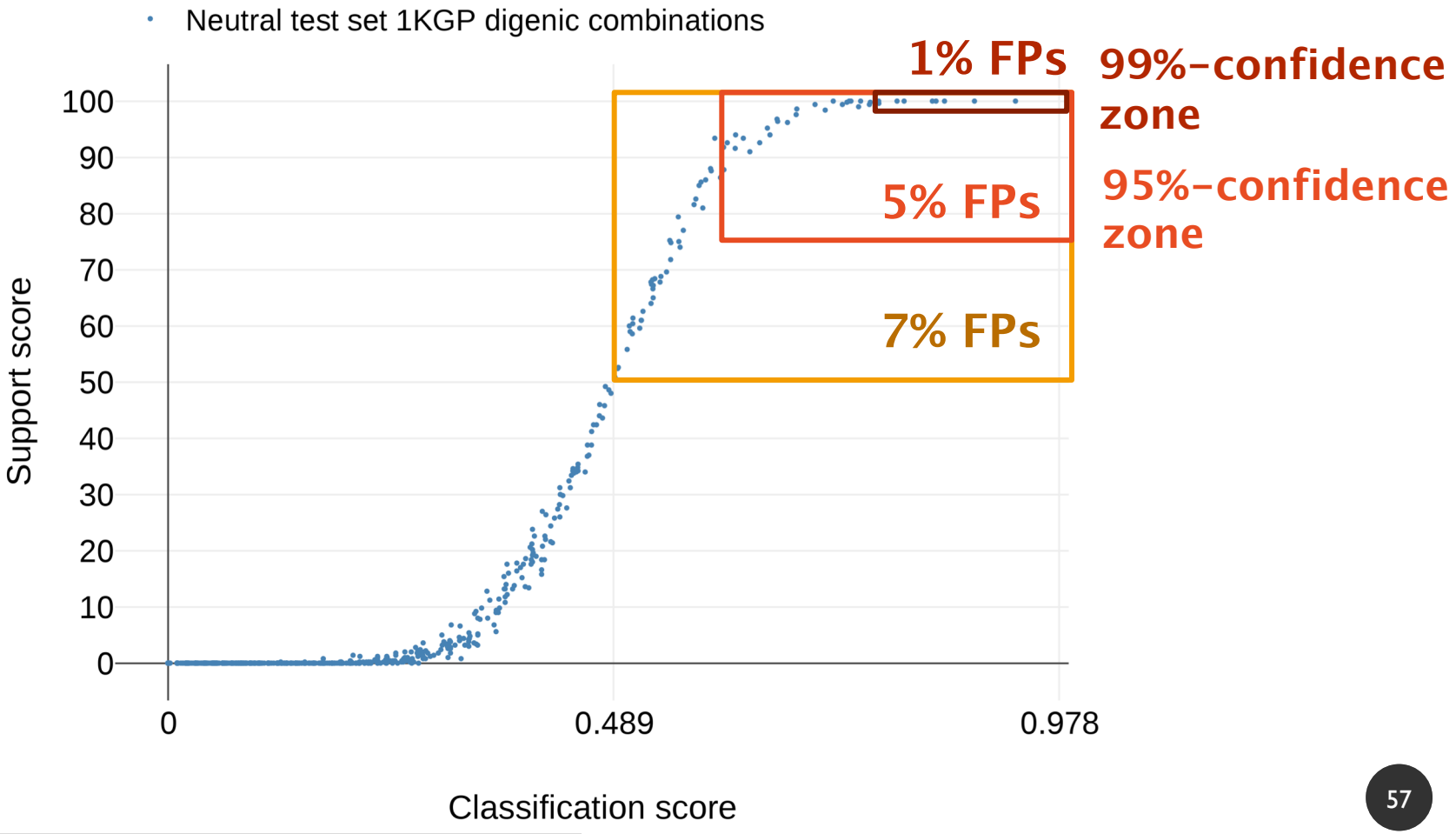
DIDA COMBINATIONS FORM AN S-PLOT



Source: Papadimitriou, *et al.* (2019). *PNAS.* **116**(24), 11878

VARCOPP: THE PATHOGENICITY PREDICTOR

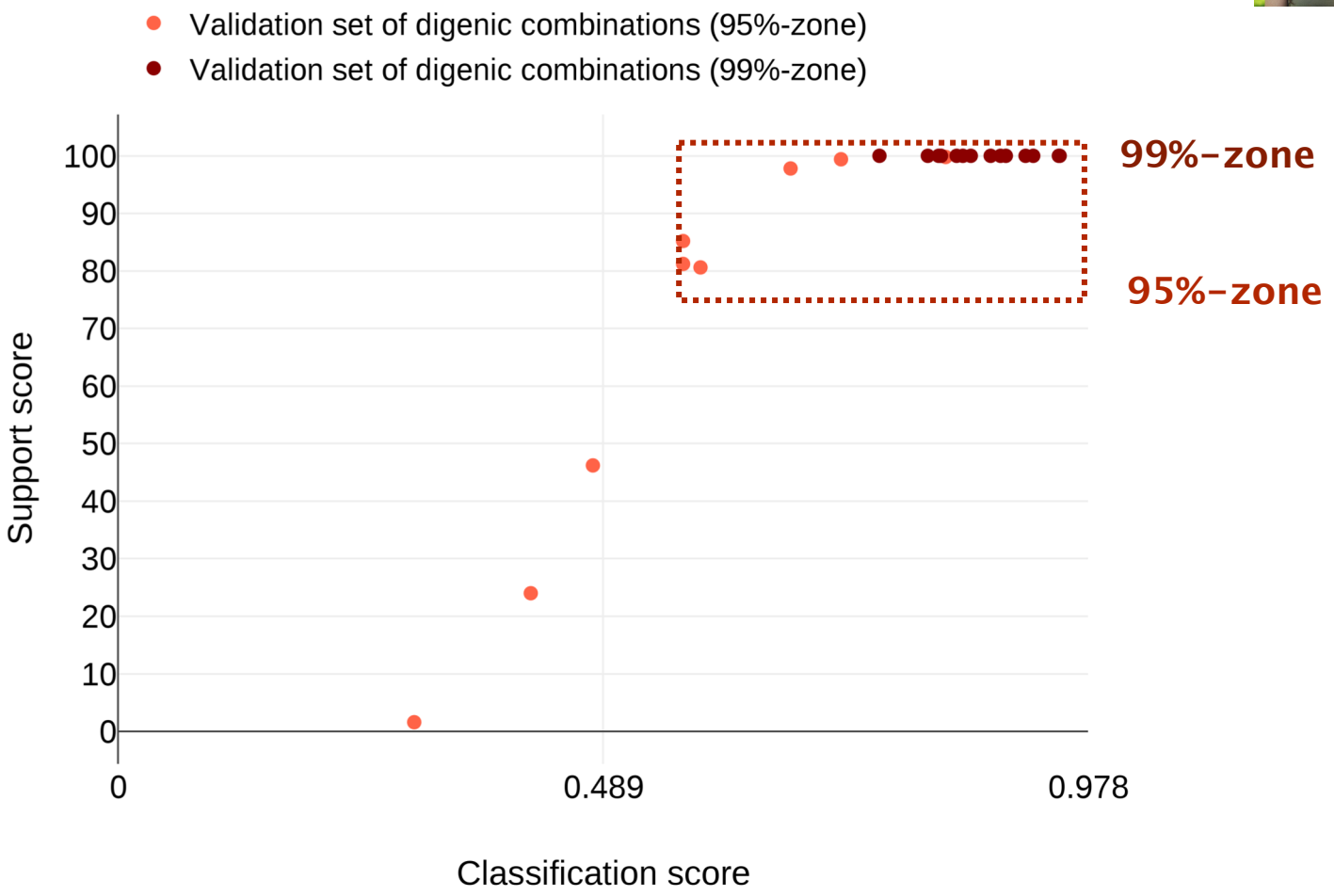
TESTING WITH UNKNOWN 1KGP DATA



Source: Papadimitriou, *et al.* (2019). *PNAS.* **116**(24), 11878

VARCOPP: THE PATHOGENICITY PREDICTOR

TESTING WITH UNKNOWN DISEASE DATA



Source: Papadimitriou, *et al.* (2019). *PNAS.* **116**(24), 11878

VARCOPP: THE PATHOGENICITY PREDICTOR

https://varcopp.ibsquare.be

Sofia Papadimitriou

Nassim Versbraegen

Submit your variants

You can either **insert each variant manually** with the six column boxes, or **copy-paste a complete variant list** directly in the white box (one variant at each line with tab or space delimited columns), or **upload a VCF file**.
Please also specify, if available, the **gender** information of the individual, as X-linked variants are handled differently between males and females.
Further information on how to upload your data is provided in the **About** page.

Sex:
Male

CHROM | POS | ID | REFERENCE | ALTERNATIVE | ZYGOSITY | +

#CHROM POSID REF ALT ZYGOSITY

Example for copy-pasted variant list:
1 69621 . A - Heterozygous
2 177054850 . C G Heterozygous
16 3254467 . CCTT C Heterozygous
X 107841975 . C A Homozygous

Vcf file:
Choose file No file chosen

Example VCF file:
Download

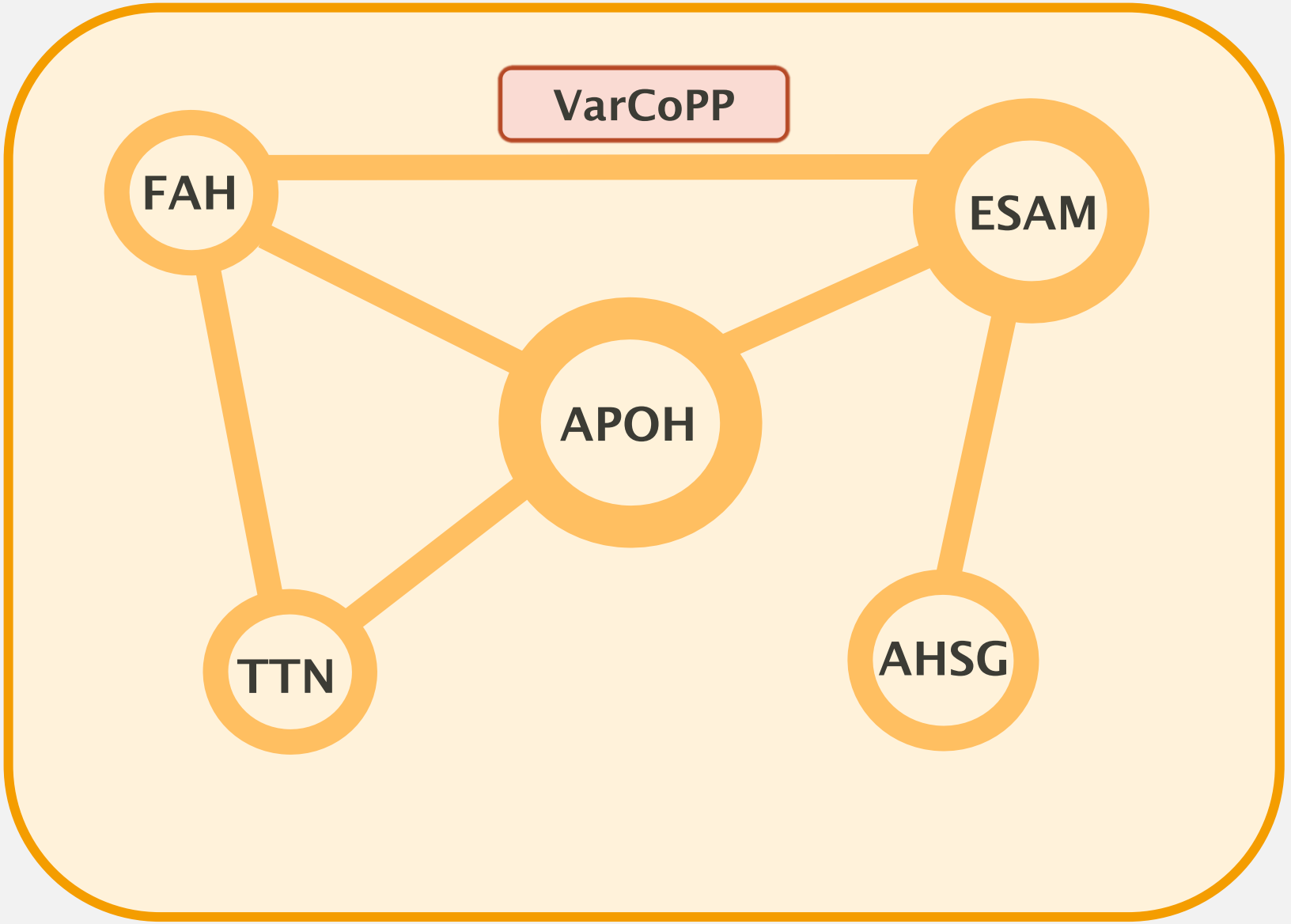
Submit

By clicking on the submit button, this page will start loading until the predictions are finished. The loading time can range between a couple of seconds to several minutes, based on the amount of data you have uploaded.

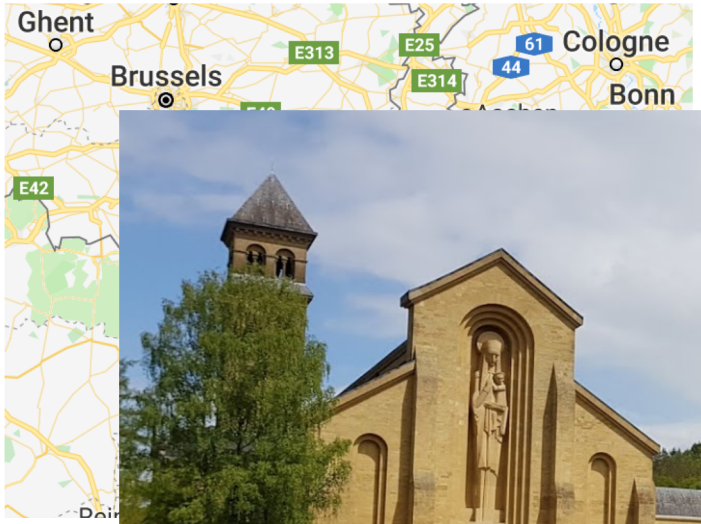
Source: Papadimitriou, *et al.* (2019). *PNAS.* **116**(24), 11878

HYPOTHESIS 3:
Using pathogenicity predictions on **gene pairs** we discover **oligogenic disease signatures** with the use of **networks**.

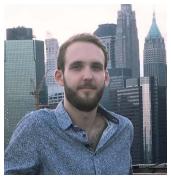
DIFFERENT APPROACH: FROM EDGES TO NODES



INSPIRED BY THE BELGIAN TRADITIONS



ORVAL: AN OLIGOGENIC ANALYSIS PLATFORM


Alexandre Renaux


Sofia Papadimitriou


Nassim Versbraegen


Charlotte Nachtegael


Simon Boutry

https://orval.ibsquare.be

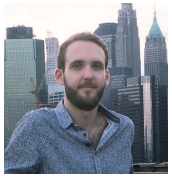
ORVAL: Oligogenic Resource for Variant AnaLysis

A platform for the prediction and exploration of candidate disease-causing oligogenic variant combinations

Run it! Learn more »

Source: Renaux, *et al.* (2019). *Nucleic Acids Research.* **47**(W1), W93

ORVAL: AN OLIGOGENIC ANALYSIS PLATFORM



Patient with disease

Genome sequencing

Variant calling

ORVAL

Submit and filter your variants

Submit a variant list of a **single individual** (VCF or tab-delimited list) and **filter** your variants based on their Minor Allele Frequency (MAF), their position in the gene and/or based on a specific gene panel of your choice.

Predict candidate pathogenic combinations

Predict candidate pathogenic combinations of variants in any gene pair with **VarCoPP** and further predict their digenic effect (True Digenic, Monogenic with a Modifier variant or Dual Diagnosis) with the **Digenic Effect Predictor**.

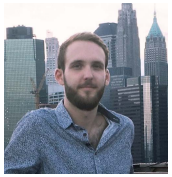
Explore oligogenic signatures

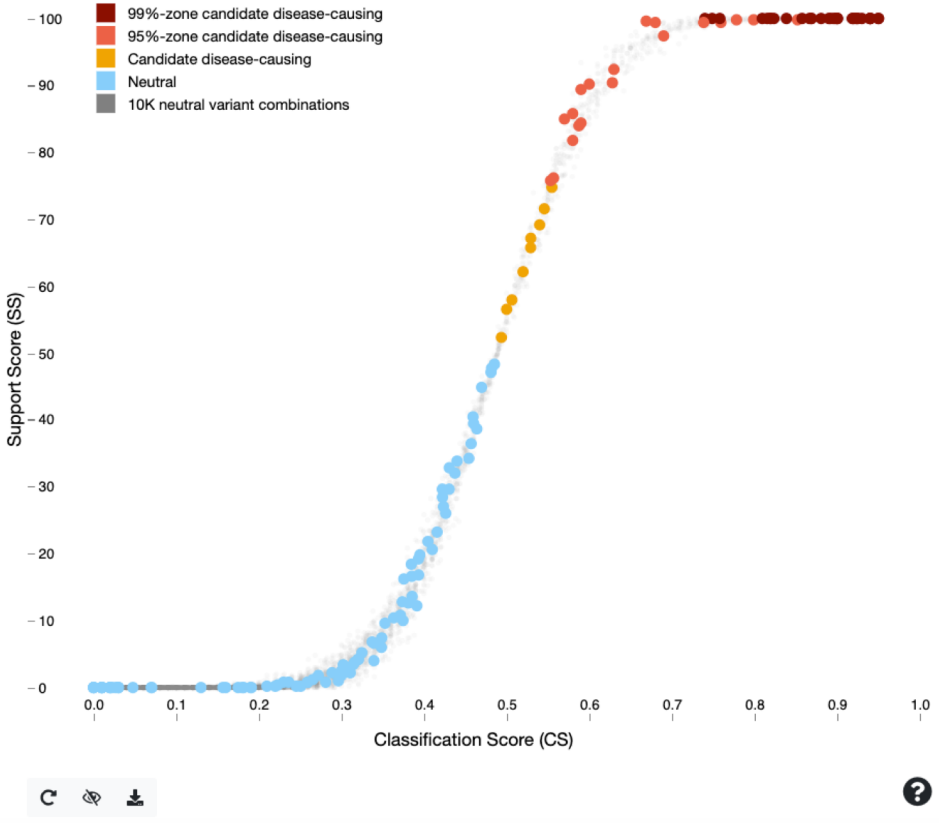
Investigate potential oligogenic disease signatures by exploring the **predicted gene networks** and examine them in the context of their pathways, protein-protein interactions and cellular locations.

Source: Renaux, *et al.* (2019). *Nucleic Acids Research.* **47**(W1), W93

ORVAL: AN OLIGOGENIC ANALYSIS PLATFORM

BI-LOCUS PATHOGENICITY PREDICTIONS





Search: Gene names/variants

Gene Pair	Variant combination (Click for more details)	VarCoPP Score	
		Classif.	Support
AHSG ESAM	3:186338564:C:T 11:124626569:G:A	0.95	100.00
AHSG FAH	3:186338564:C:T 15:80472526:C:T	0.95	100.00
VEGFC AHSG	4:177650866:C:T 3:186338564:C:T	0.94	100.00
AHSG SERPINA1	3:186338564:C:T 14:94844947:C:T	0.94	100.00
AHSG APOH	3:186338564:C:T 17:64210580:A:G	0.94	100.00
VEGFC FAH	4:177650866:C:T 15:80472526:C:T	0.94	100.00
AHSG POSTN	3:186338564:C:T 13:38156538:C:T	0.93	100.00
VEGFC APOH	4:177650866:C:T 17:64210580:A:G	0.93	100.00
VEGFC AHSG	4:177650866:C:T 3:186338397:C:A	0.92	100.00
SERPINA1 APOH	14:94844947:C:T 17:64210580:A:G	0.92	100.00

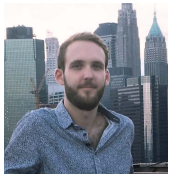
1-10 / 155 variant pairs

Previous 1 2 3 4 5 … 16 Next

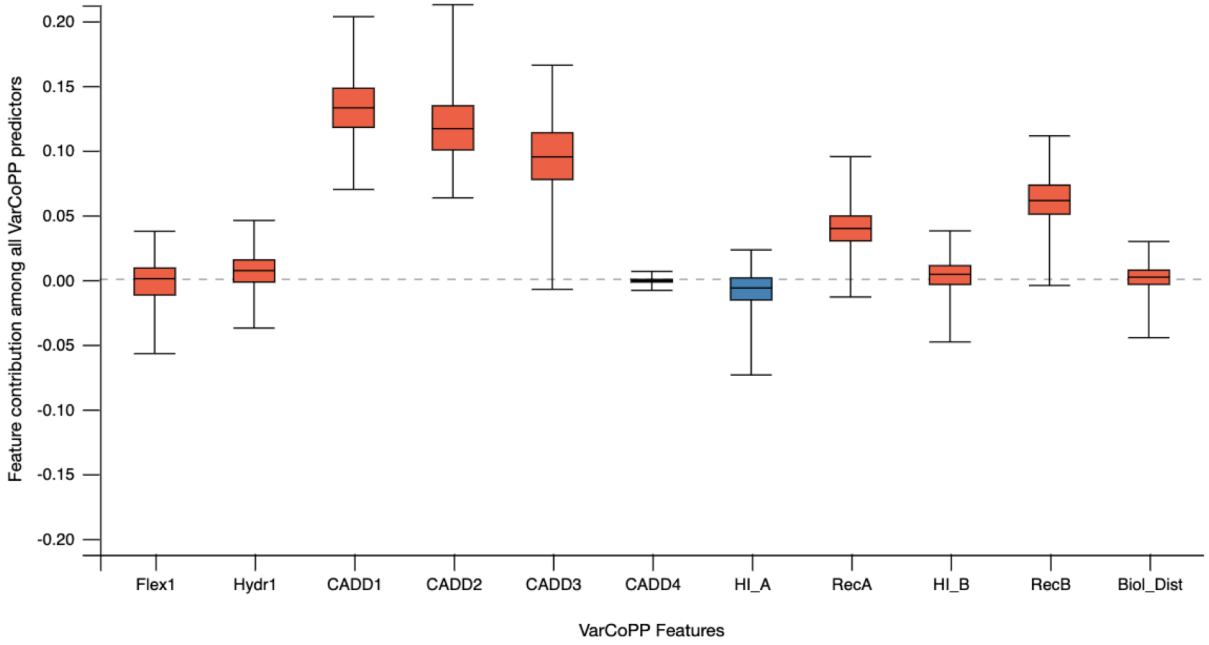
Source: Renaux, *et al.* (2019). *Nucleic Acids Research.* **47**(W1), W93

ORVAL: AN OLIGOGENIC ANALYSIS PLATFORM

BI-LOCUS PATHOGENICITY PREDICTIONS



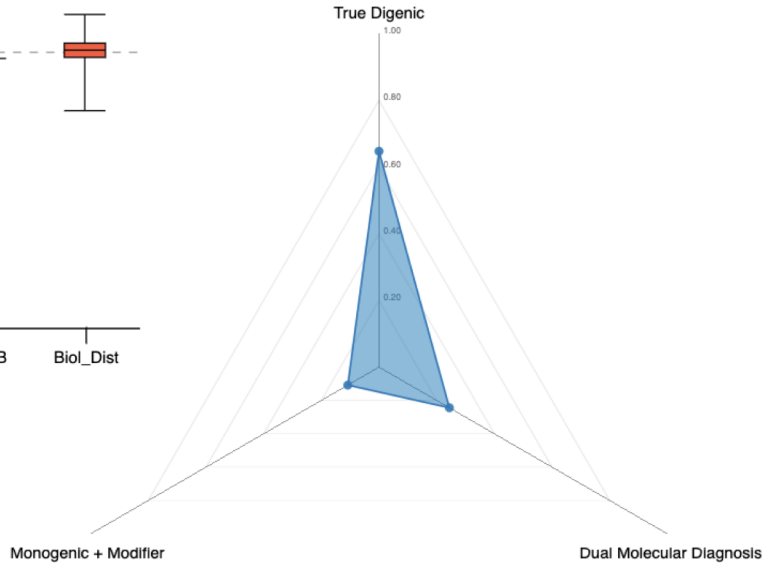
Pathogenicity prediction: feature contributions



Preference for the disease-causing class

Preference for the neutral class

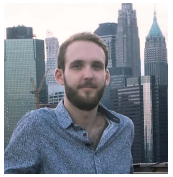
Digenic Effect Prediction



Source: Renaux, *et al.* (2019). *Nucleic Acids Research.* **47**(W1), W93

ORVAL: AN OLIGOGENIC ANALYSIS PLATFORM

FURTHER EXTERNAL ANNOTATIONS



Annotations

	AHSG	FAH
CHROMOSOME	3	15
GENE NAME	AHSG	FAH
ENSEMBL GENE ID	ENSG00000145192	ENSG00000103876
UNIPROT PROTEIN ID	P02765	P16930
GDI	105.08096	149.7435
P(HAPLOINSUFFICIENCY)	0.47188	0.1414
P(RECESSIVENESS)	0.39186	0.24341

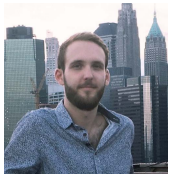
	Variant chr3:186338564 C>T	Variant chr15:80472526 C>T
GENOMIC CHANGE	g.186338564C>T	g.80472526C>T
cDNA CHANGE	c.949C>T	c.1021C>T
PROTEIN CHANGE	p.R317C	p.R341W
ZYGOSITY	Homozygous	Homozygous
RS ID	rs35457250	rs11555096
ENSEMBL TRANSCRIPT ID	ENST00000411641	ENST00000561421
CADD	6.093498	6.361174
ExAC ALLELE FREQ.	0.01085	0.021979

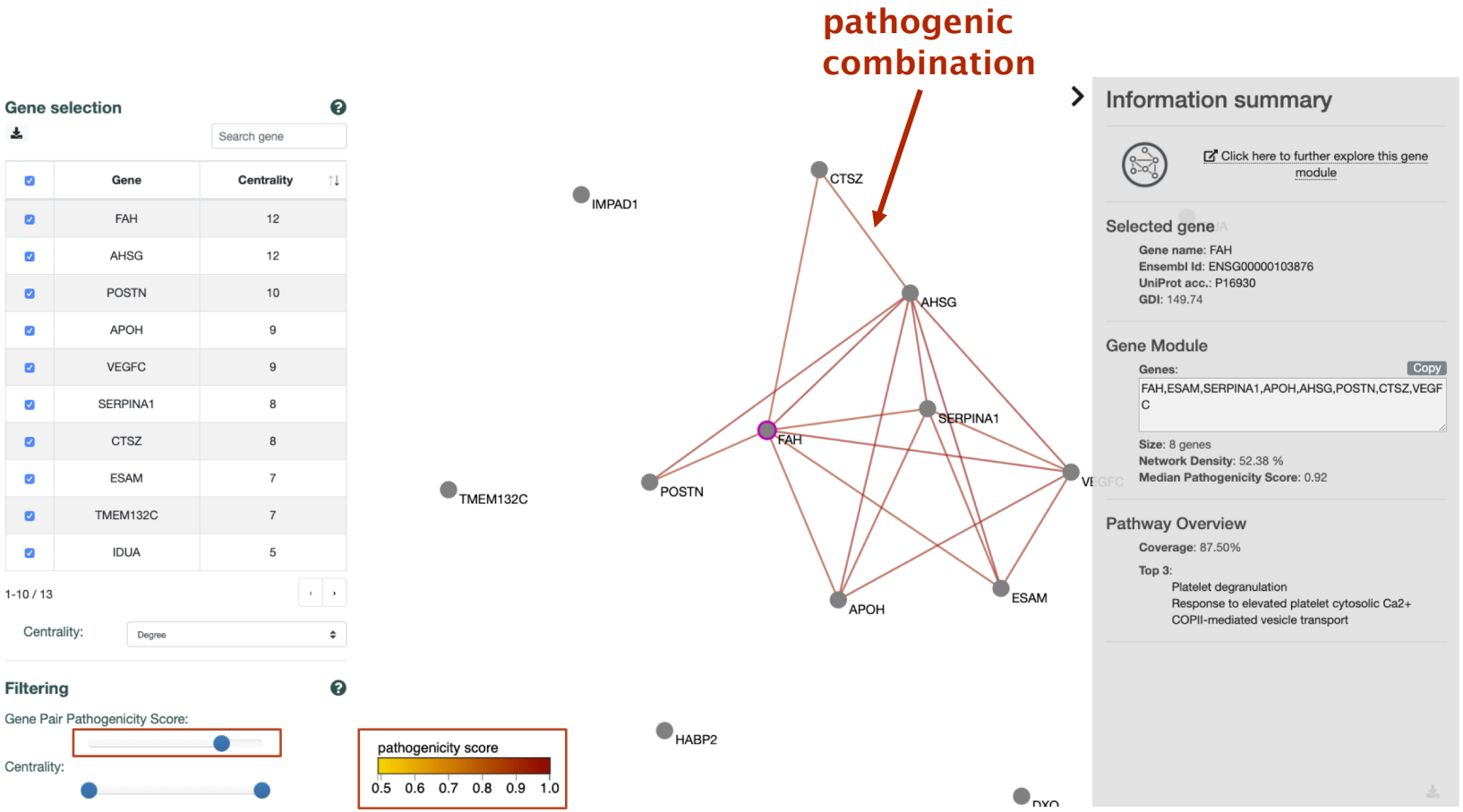
	{ AHSG - FAH }
BIOLOGICAL DISTANCE	14.602

Source: Renaux, *et al.* (2019). *Nucleic Acids Research.* **47**(W1), W93

ORVAL: AN OLIGOGENIC ANALYSIS PLATFORM

PREDICTED PATHOGENIC NETWORKS

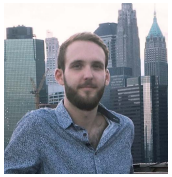




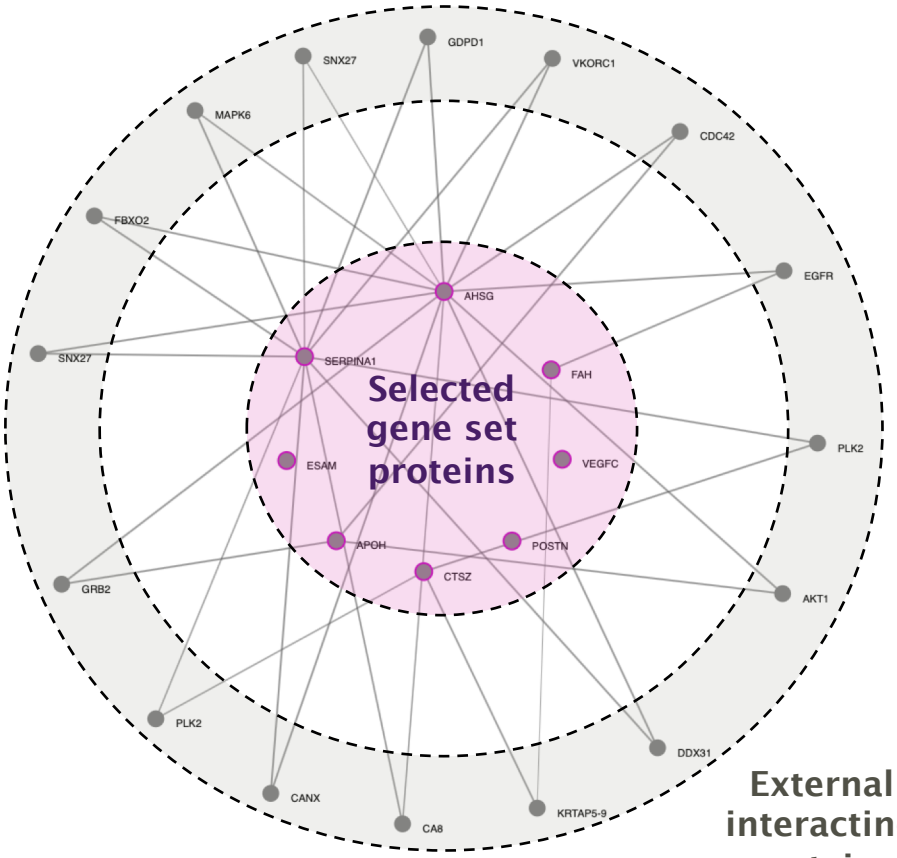
Source: Renaux, *et al.* (2019). *Nucleic Acids Research.* **47**(W1), W93

ORVAL: AN OLIGOGENIC ANALYSIS PLATFORM

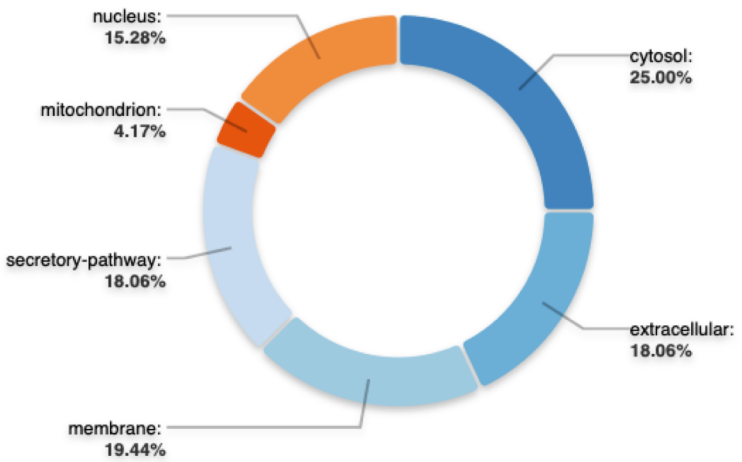
NETWORK MODULE INFORMATION



Protein–protein interactions



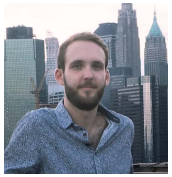
Cellular compartment location



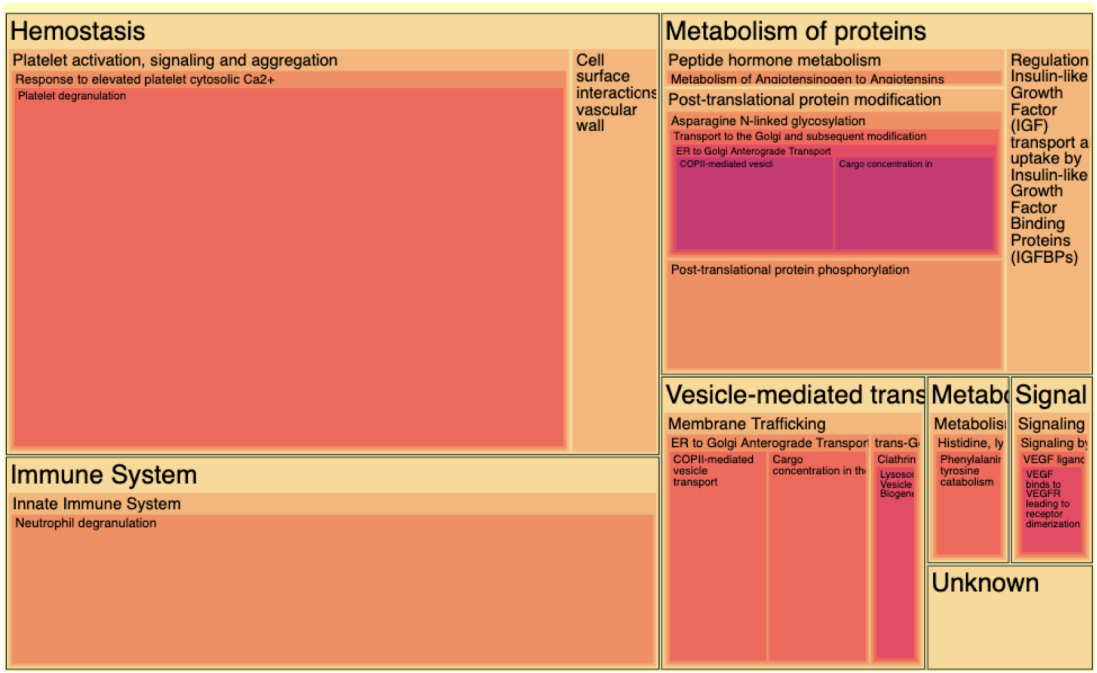
Source: Renaux, *et al.* (2019). *Nucleic Acids Research.* **47**(W1), W93

ORVAL: AN OLIGOGENIC ANALYSIS PLATFORM

NETWORK MODULE INFORMATION



Pathway TreeMap



Pathway – Gene Mappings

Search: Gene / Pathway

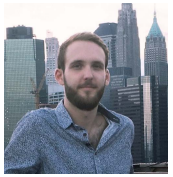
Pathway	Genes
Hemostasis	ESAM,VEGFC,AHSG,SERPINA1,APOH
Platelet activation, signaling and aggregation	AHSG,VEGFC,APOH,SERPINA1
Response to elevated platelet cytosolic Ca2+	AHSG,VEGFC,APOH,SERPINA1
Platelet degranulation	AHSG,VEGFC,APOH,SERPINA1
Immune System	SERPINA1,AHSG,CTSZ
Innate Immune System	SERPINA1,AHSG,CTSZ
Neutrophil degranulation	SERPINA1,AHSG,CTSZ
Metabolism of proteins	SERPINA1,AHSG,CTSZ
Post-translational protein modification	SERPINA1,AHSG,CTSZ
Asparagine N-linked glycosylation	SERPINA1,CTSZ

1-10 / 37

Source: Renaux, *et al.* (2019). *Nucleic Acids Research.* **47**(W1), W93

ORVAL: AN OLIGOGENIC ANALYSIS PLATFORM

EXPLORING A REAL PUBLISHED CASE



Recent clinical study of a patient with **mild hypertrophic cardiomyopathy**.

> *We report the case of a master athlete carrying **trigenic mutations** in desmoglein-2 (**DSG2**), desmocollin-2 (**DSC2**) and heavy chain myosin 6 (**MYH6**) (…).*

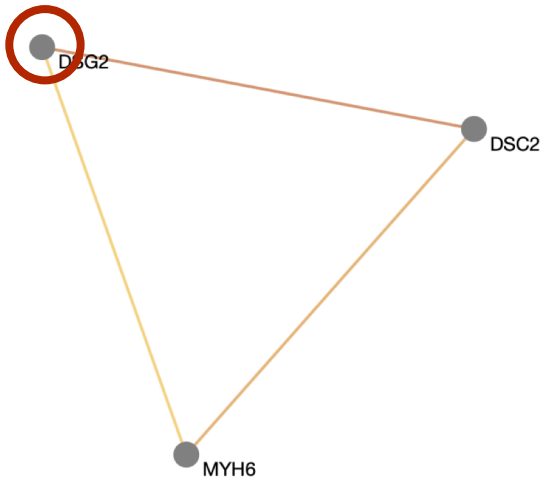
J Electrocardiol. 2019 Mar - Apr;53:95-99. doi: 10.1016/j.jelectrocard.2019.01.002. Epub 2019 Jan 2.

Sudden death in mild hypertrophic cardiomyopathy with compound DSG2/DSC2/MYH6 mutations: Revisiting phenotype after genetic assessment in a master runner athlete.

Castellana S[1], Mastroianno S[2], Palumbo P[3], Palumbo O[3], Biagini T[1], Leone MP[3], De Luca G[2], Potenza DR[2], Amico CM[2], Mazza T[1], Russo A[2], Di Stolfo G[2], Carella M[4].

Pathogenic network of the patient

homozygous

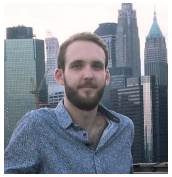


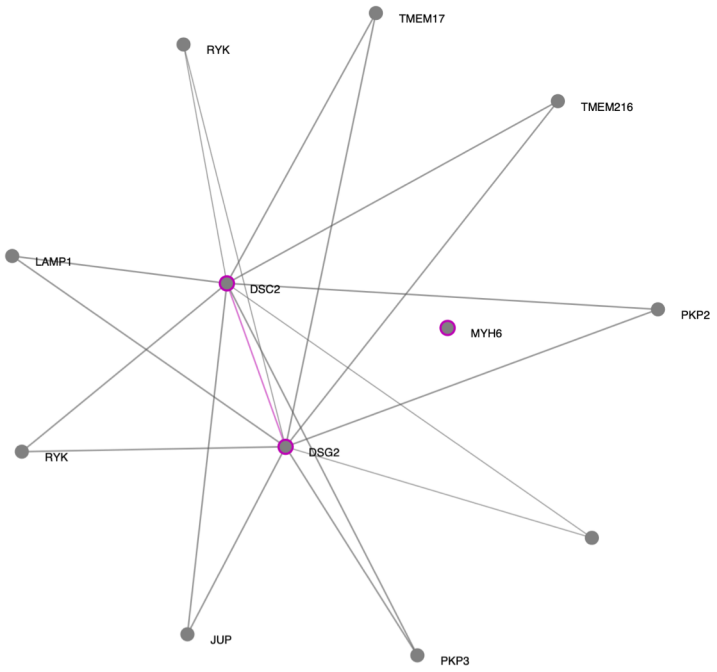
Gene Pair	Variant combination (Click for more details)	VarCoPP Score	
		Classif. ↑↓	Support ↑↓
DSC2 DSG2	18:28648000:-:TC 18:29104714:A:G	0.81	100.00
DSC2 MYH6	18:28648000:-:TC 14:23865901:A:G	0.69	99.40
DSG2 MYH6	18:29104714:A:G 14:23865901:A:G	0.57	85.60

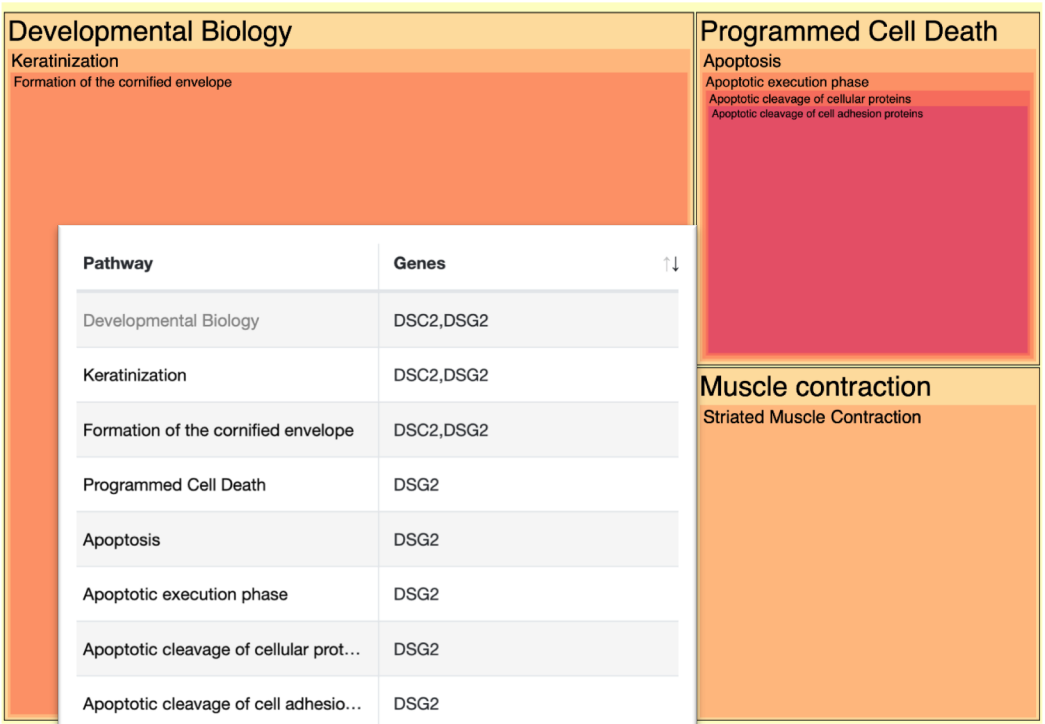
Source: Renaux, *et al.* (2019). *Nucleic Acids Research.* **47**(W1), W93

ORVAL: AN OLIGOGENIC ANALYSIS PLATFORM

EXPLORING A REAL PUBLISHED CASE







Pathway	Genes
Developmental Biology	DSC2,DSG2
Keratinization	DSC2,DSG2
Formation of the cornified envelope	DSC2,DSG2
Programmed Cell Death	DSG2
Apoptosis	DSG2
Apoptotic execution phase	DSG2
Apoptotic cleavage of cellular prot…	DSG2
Apoptotic cleavage of cell adhesio…	DSG2
Muscle contraction	MYH6
Striated Muscle Contraction	MYH6

Source: Renaux, *et al.* (2019). *Nucleic Acids Research.* **47**(W1), W93

CONCLUSIONS

- Most tools and techniques have focused on a univariate analysis towards oligogenic diseases
- We appear to have good quality pathogenicity predictions for variant combinations in pairs of genes
- We show the usefulness of interpretable machine learning methods in medical genetics
- We can distill oligogenic modules, which could provide a new way of exploring a patient's exome

THERE IS STILL A LOT OF WORK TO DO

- Validating our tools on multiple cohorts
(neurodevelopmental diseases, deafness, epilepsy,
Brugada syndrome, congenital heart defects…)

- Provide trio analysis in ORVAL
- Use phenotypic information for filtering/annotation
- New features
- Create disease-specific predictors
- Transform DIDA into a community effort

ACKNOWLEDGING THE OLIGOGENIC TEAM

Our professors and senior researchers

Tom Lenaerts

Ann Nowé

Guillaume Smits

Sonia van Dooren

Jan Aerts

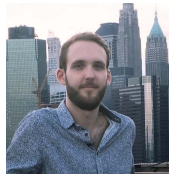
Yves Moreau

Our PhD / Master students, junior researchers

Sofia Papadimitriou

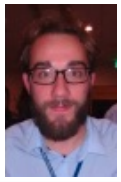
Nassim Versbraegen

Charlotte Nachtegael



Alexandre Renaux

Simon Boutry



Andrea Gazzo

Dorien Daneels

Claudio Reggiani

Arnau Dillen

Aziz Fouché