

Uncovering novel genetic factors of pediatric sepsis through new tools for **precision genomic medicine.**

Dylan Lawless
Pauline Rogg
Manon Bouzereau
Robin Fallegger
Zaira Seferbekovatbd
Sina Ruger
Shweta Pipaliya
Valeriia Timonina
Konstantin Pipadin
Ali Saadat
Zhi Ming Xu
Christian Thorball
Flavia Hodell
Philipp Agyeman
Jacques Fellay
Luregn J. Schlapbach
for the Swiss Pediatric Sepsis Study

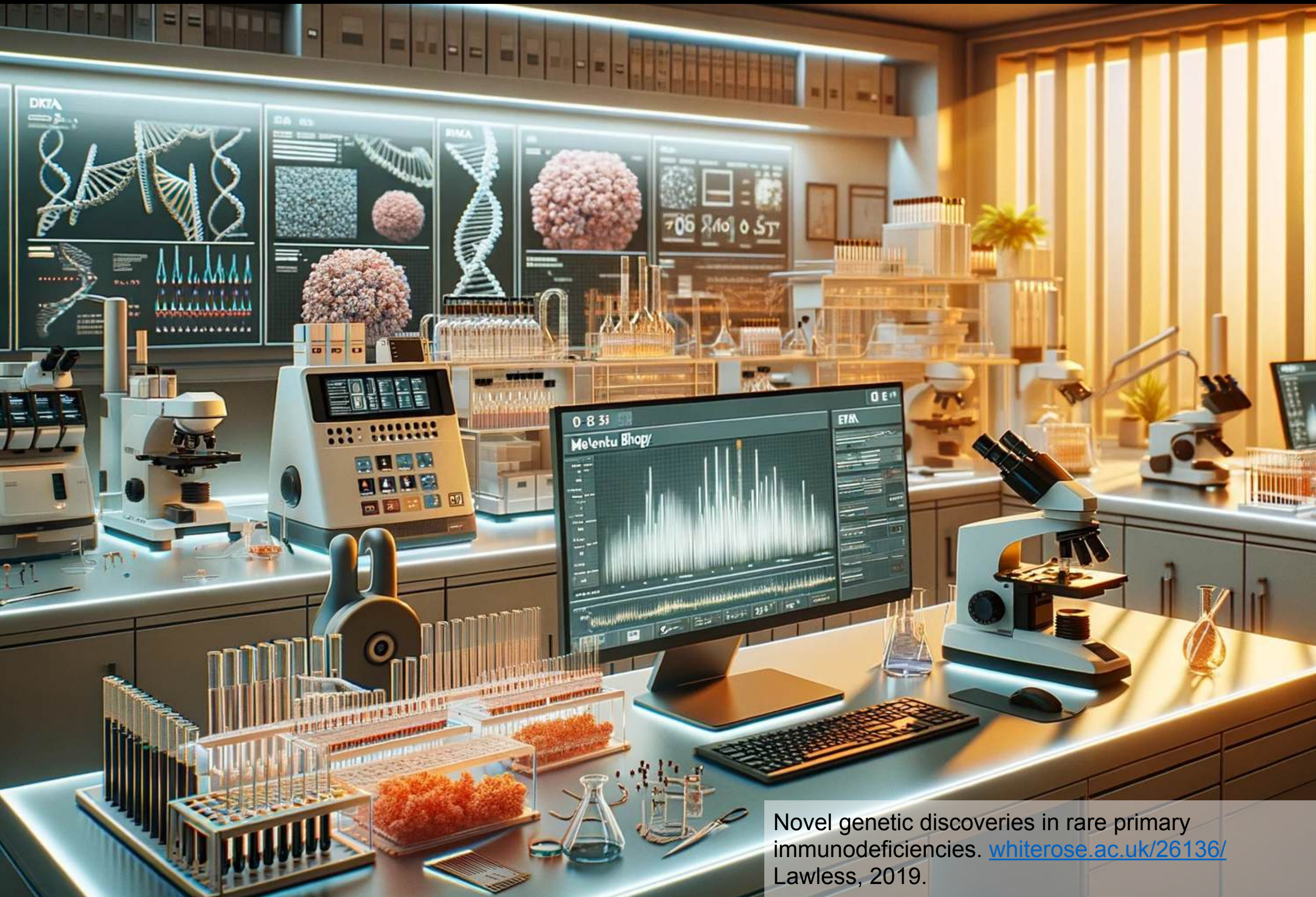
**“Learn the facts, Steed-Asprey
used to say, then try on the
stories like clothes.”**

Tinker, Tailor, Soldier, Spy.
John le Carré

A face behind the research



Decoding the puzzle



Novel genetic discoveries in rare primary immunodeficiencies. whiterose.ac.uk/26136/ Lawless, 2019.

Journey's verdict



TET2 Publication

DNA sequencing technology

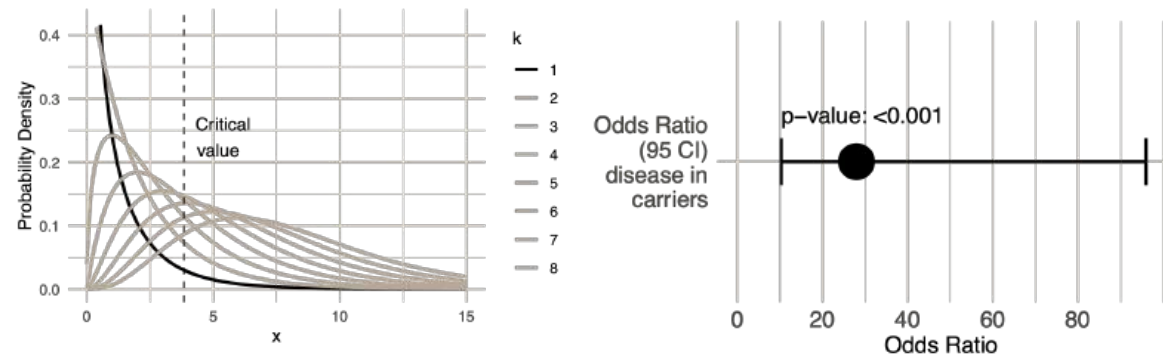
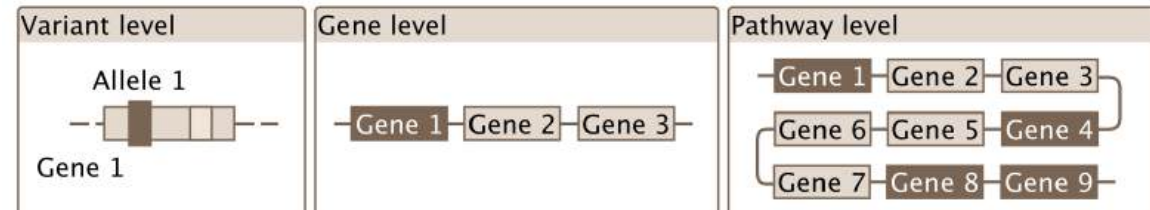
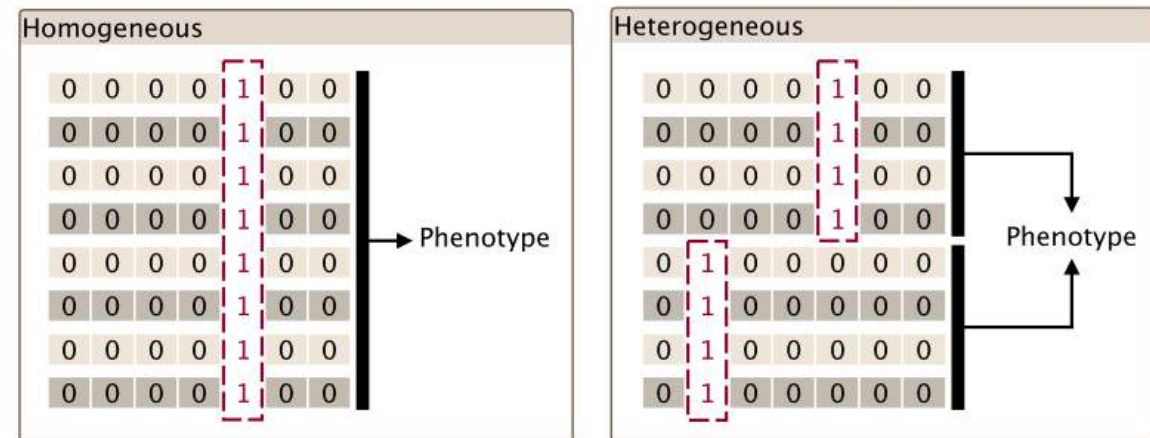
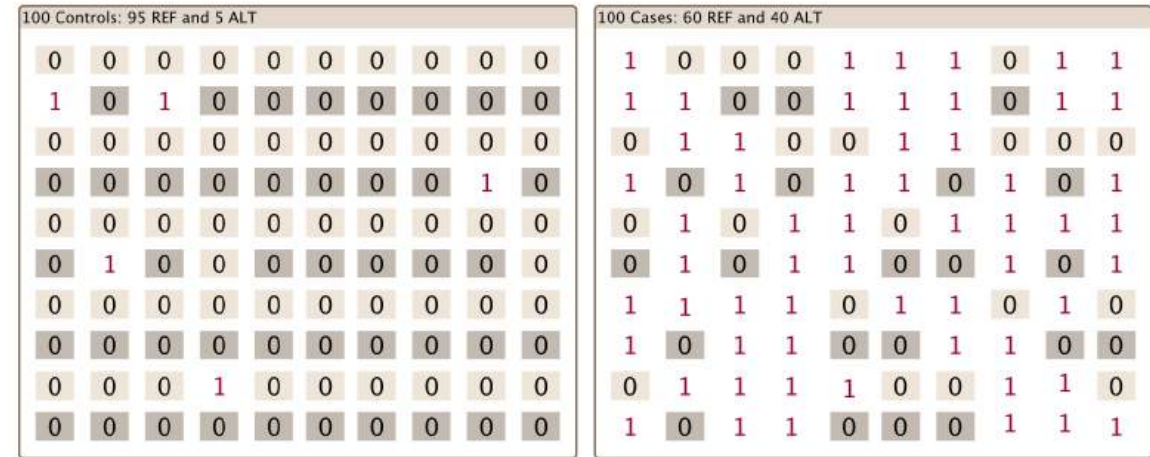
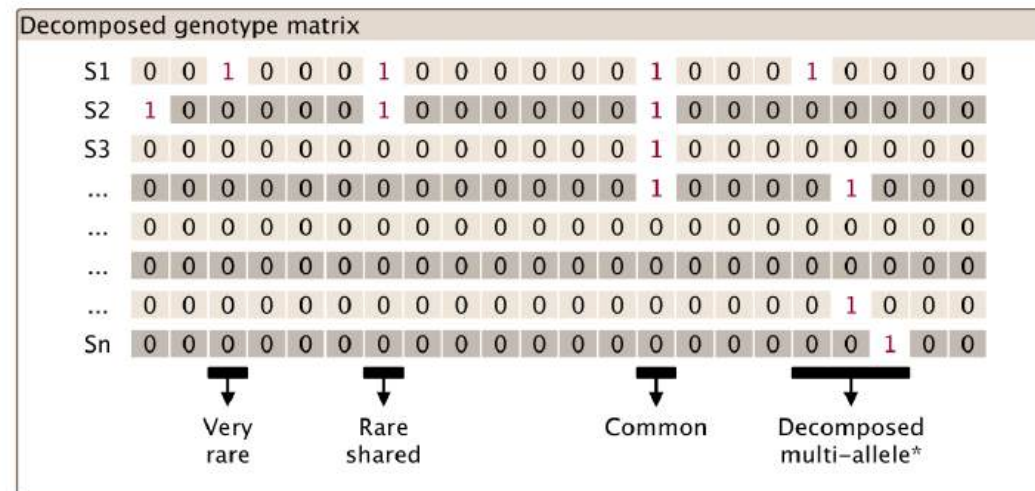
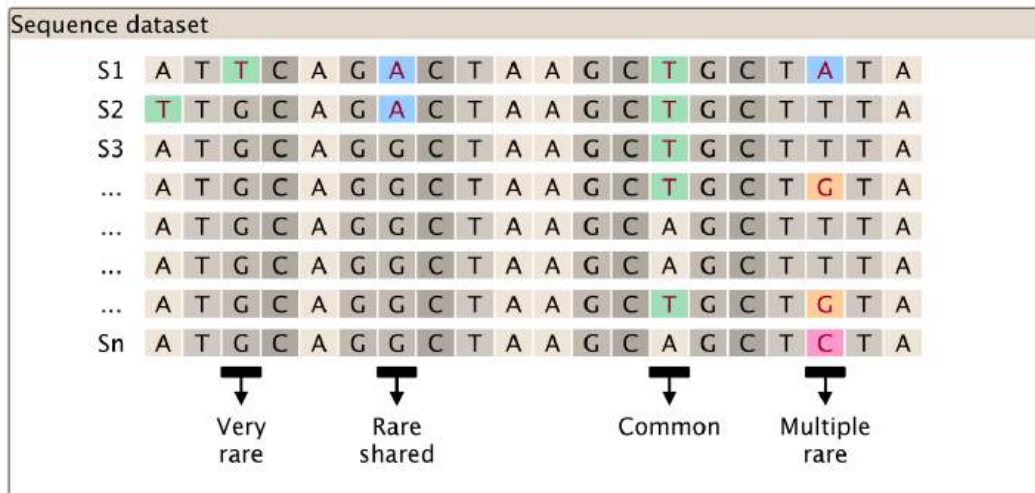
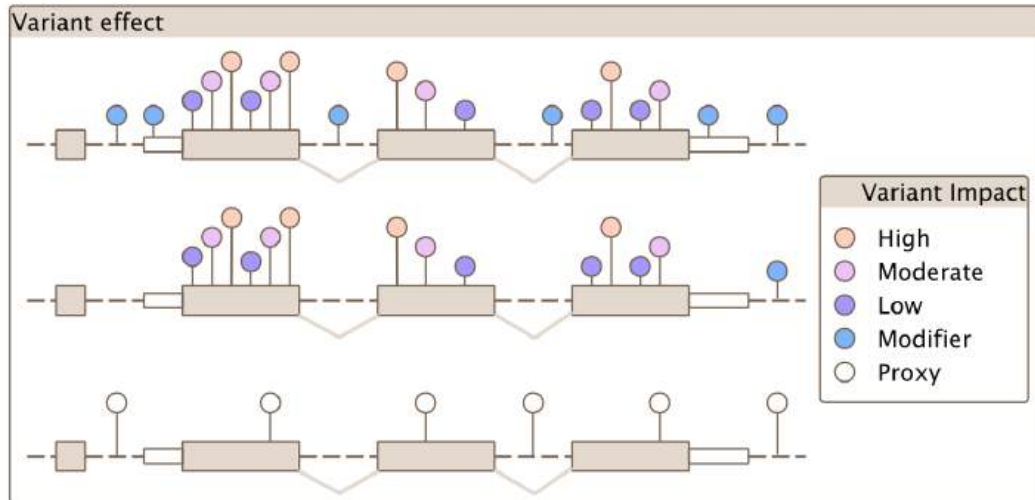


JACI. The Ever-Increasing Array of Novel Inborn Errors of Immunity: an Interim Update by the IUIS Committee [10.1007/s10875-021-00980-1](https://doi.org/10.1007/s10875-021-00980-1) Tangye et al., 2021.

Blood. Germline TET2 Loss-Of-Function Causes Childhood Immunodeficiency and Lymphoma. [10.1182/blood.2020005844](https://doi.org/10.1182/blood.2020005844) Jarmila Stremenova Spegarova, and Dylan Lawless, et. al.

Scale

From wet to dry



From wet to dry



From wet to dry

Variant annotation

Variant	Set	Sample	Genotype	Age	Sex	Cohort AF	CADD phred	REVEL	ClinVar	FATHMM	OMIM	PANTHER	Gene ontology	...	GnomAD AF
1	A	S1	0	18	0	.3	25.90	.902	PL	.7	179615	11539	0002331	...	3E-05
1	A	...	0	21	1	.3	25.90	.902	PL	.7	179615	11539	0002331	...	3E-05
1	A	Sn	1	45	0	.3	25.90	.902	PL	.7	179615	11539	0002331	...	3E-05
2	A	S1	1	18	0	.01	29.3	.783	P	.9	179615	11539	0002331	...	7E-06
2	A	...	0	21	1	.01	29.3	.783	P	.9	179615	11539	0002331	...	7E-06
2	A	Sn	0	45	0	.01	29.3	.783	P	.9	179615	11539	0002331	...	7E-06
3	B	S1	1	18	0	.5	25.9	.888	PL	NA	147660	10960	0002331	...	3E-05
3	B	...	0	21	1	.5	25.9	.888	PL	NA	147660	10960	0002331	...	3E-05
3	B	Sn	0	45	0	.5	25.9	.888	PL	NA	147660	10960	0002331	...	3E-05
4	B	S1	0	18	0	.02	12.1	NA	NA	NA	147660	10960	0002331	...	3E-05
4	B	...	0	21	1	.02	12.1	NA	NA	NA	147660	10960	0002331	...	3E-05
4	B	Sn	0	45	0	.02	12.1	NA	NA	NA	147660	10960	0002331	...	3E-05

Set
level

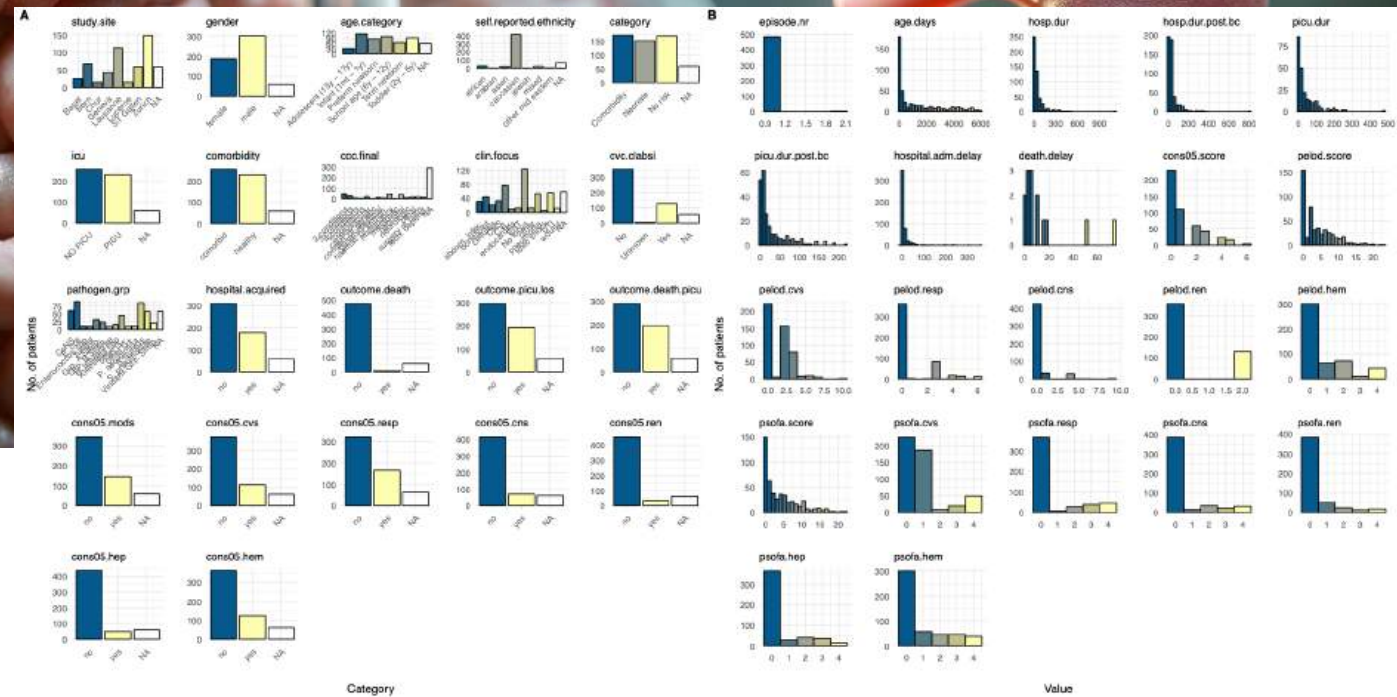
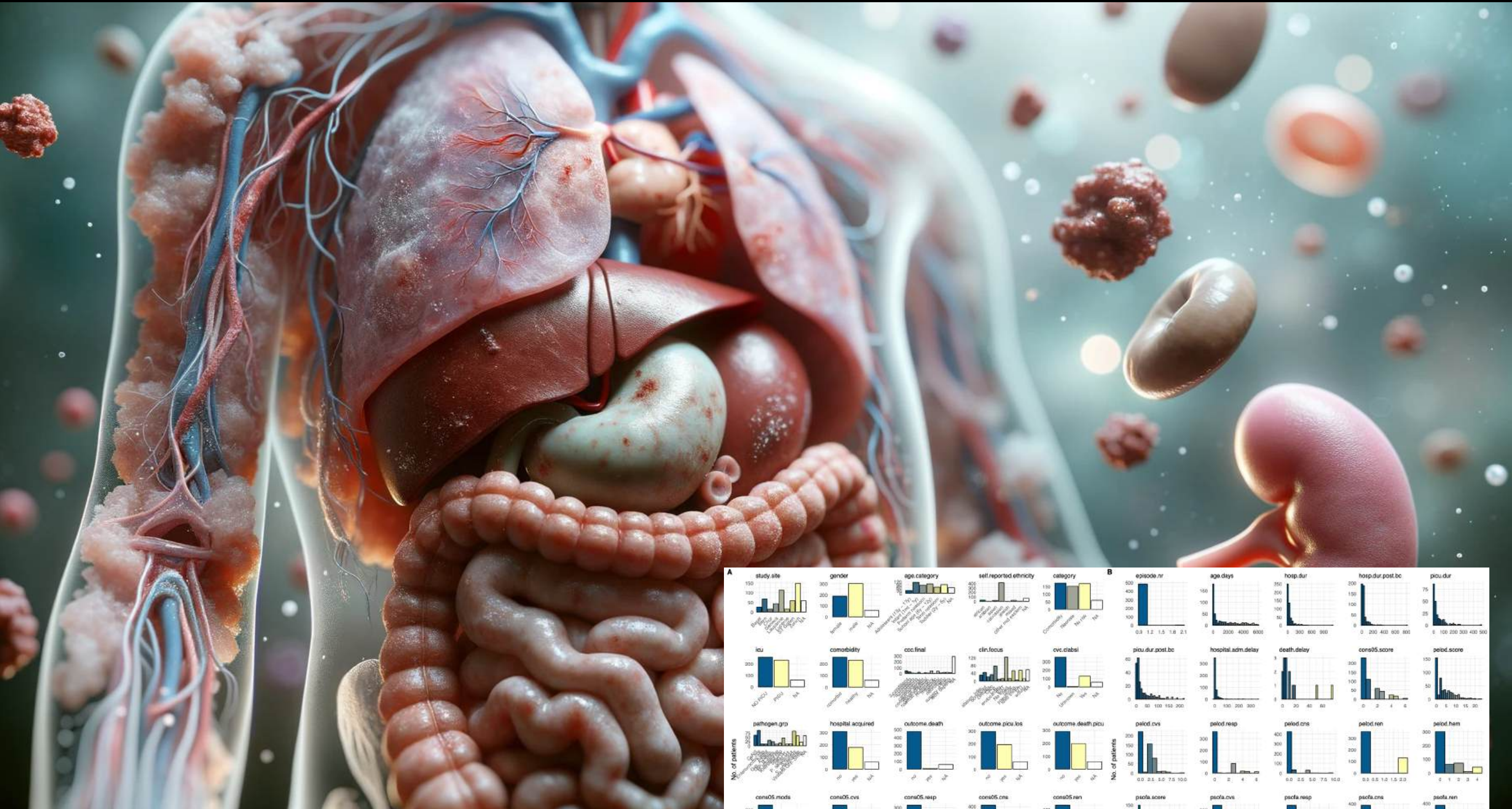
Sample
level

Variant
level

Gene
level

Ontology
level

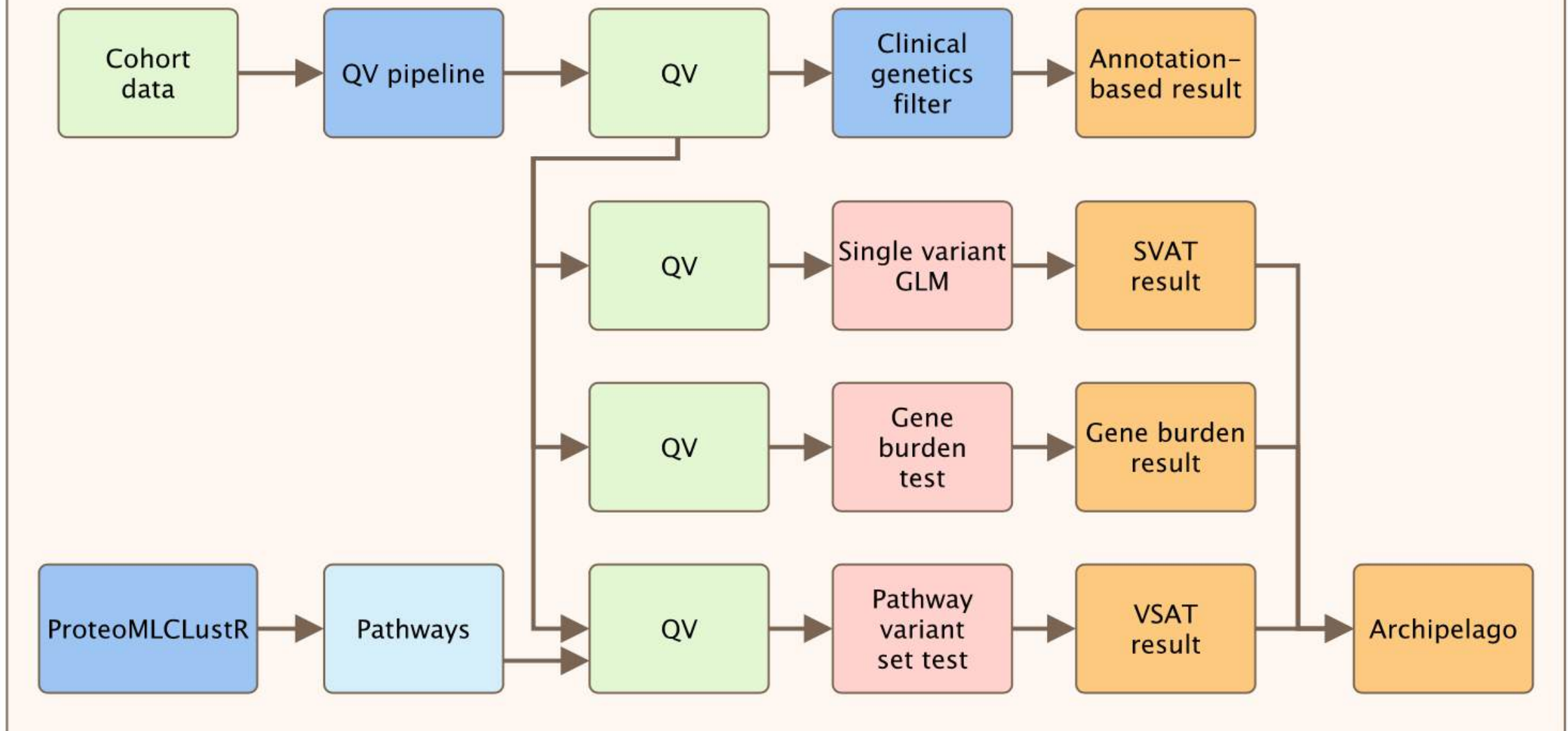
Sepsis



940 patients
548 cases
392 controls

From dot to network

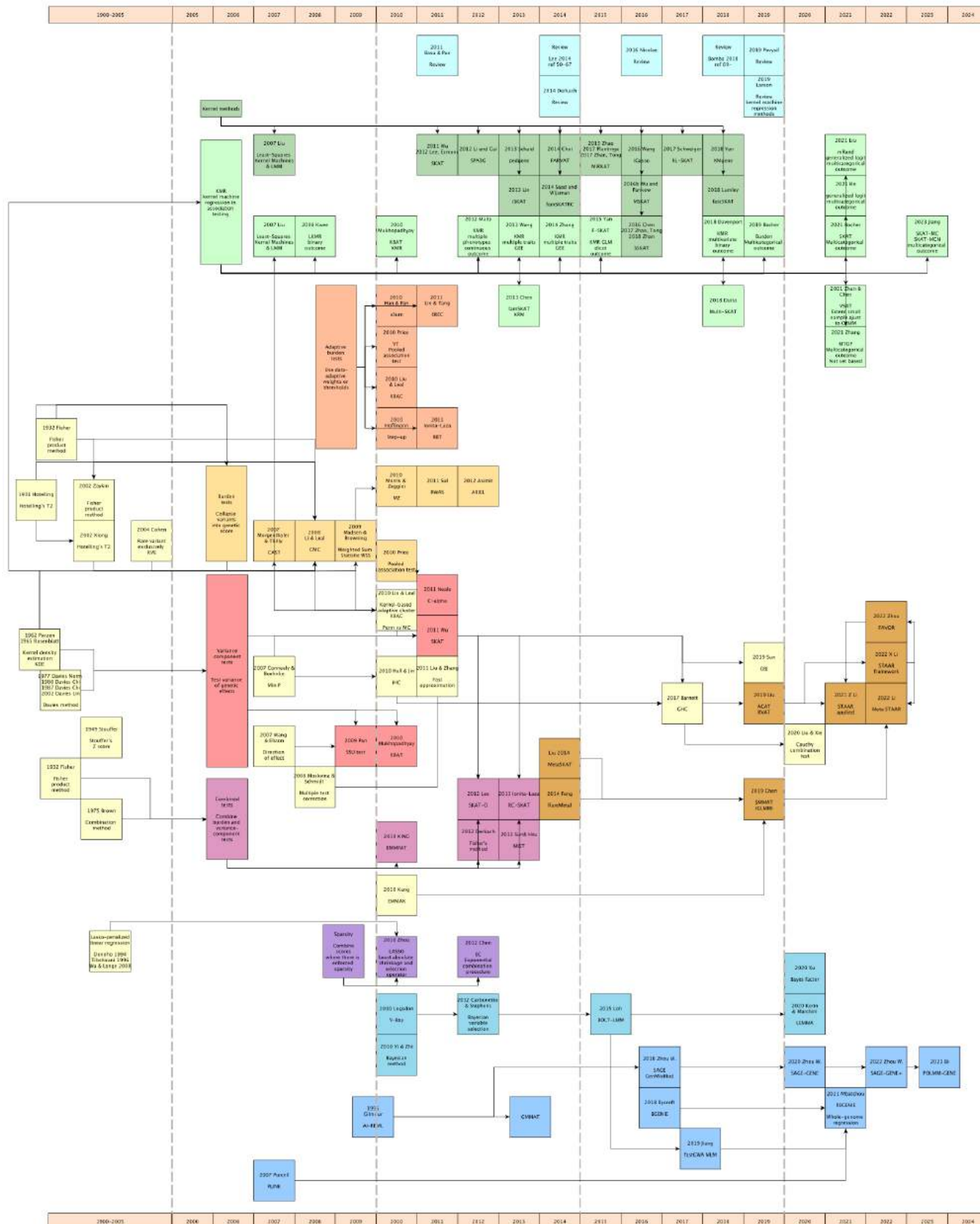
Sequence analysis and association testing overview



Collapse



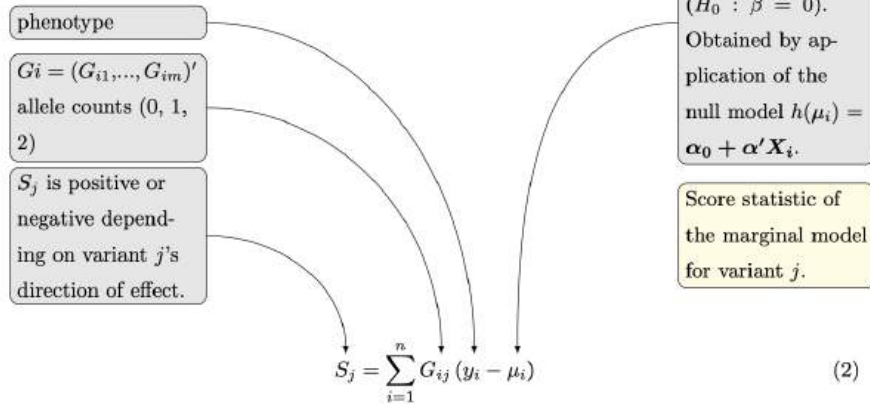
Variant analysis history



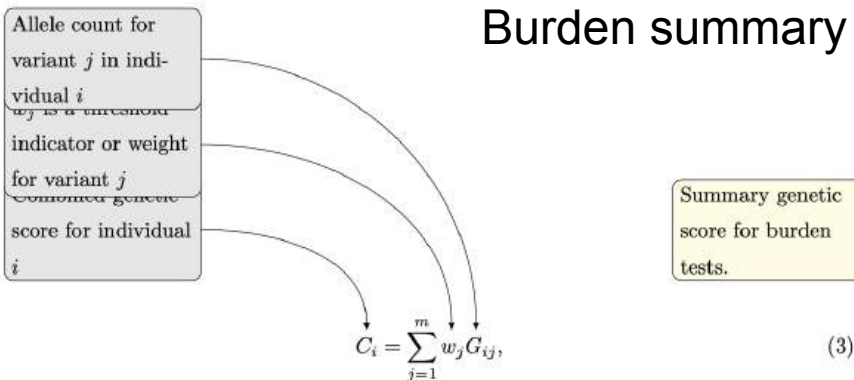
bib	Reference	method	Category_sir	Category	Description
\cite{1931hotellingGe}	1931 Hotelling (1992)	Hotelling's T2	Reference me	Statistical tes	Introduces the concept of
\cite{1932fisherStatis}	1932 Fisher (1992 re	Fisher product	Reference me	NA	NA
\cite{1949stoufferAm}	1949 Stouffer	Stouffer's Z so	Reference me	NA	NA
\cite{1956rosenblattF}	1956 Rosenblatt	Kernel density	Kernel	Kernel densit	Builds upon the previous
\cite{1975brown400}	1975 Brown	Combination n	Reference me	Meta-analysis	Introduces the Brown-F
\cite{1977daviesHypc}	1977 Davies Norm	Davies method	Reference me	Hypothesis te	discusses methods for hy
\cite{1979neymanC}	1979 Neyman	C-alpha	Reference methods		
\cite{1980daviesAlgoi}	1980 Davies Chi	Davies method	Reference me	Distribution t	derives the distribution
\cite{1987daviesHypc}	1987 Davies Chi	Davies method	Reference me	Hypothesis te	is a continuation of Davi
\cite{1995gilmourAve}	1995 Gilmour	GMMAT	Reference me	Linear mixed i	Introduces the average
\cite{1999devlinGeno}	1999 Devlin	Genomic contr	Reference me	Genomic cont	Introduces the genomic
\cite{1kg2012integral}	2012 1KG	1kG	Reference me	Human genon	Presents an integrated i
\cite{2002daviesHypc}	2002 Davies Lin	Davies method	Reference me	Hypothesis te	is a further continuation
\cite{2002xiongGener}	2002 Xiong	Hotelling's T2	Reference me	Genome asso	Introduces the generaliz
\cite{2002zaykinTrun}	2002 Zaykin	Fisher product	Reference me	Meta-analysis	Introduces the truncate
\cite{2004cohenMulti}	2004 Cohen	Rare variant e	Reference me	Rare alleles (s	A combination of metho
\cite{2005whitlockCo}	2005 Whitlock	Weighted Z	Reference me	Meta-analysis	Introduces the weighte
\cite{2006pricePincipi}	2006 Price	PCA	Reference me	Population st	Describes the use of pri
\cite{2007conneelyM}	2007 Conneely & Bo	Min P	Reference me	Multiple testi	presents a rapid metho
\cite{2007liuSemipar}	2007 Liu	Least-Squares	Kernel	Kernel metho	presents a method for si
\cite{2007morgenthal}	2007 Morgenthaler	CAST	Burden test	Genetic assoc	presents the cohort alle
\cite{2007purcellPlink}	2007 Purcell	Plink	Software imp	Genome-wide	Presents PLINK, a popul
\cite{2007wangImpro}	2007 Wang & Elston	Direction of ef	Adaptive bur	Linkage diseq	presents a weighted scc
\cite{2008hofmannKe}	2008 Hofmann	Kernel	Kernel	Kernel metho	provides an overview of
\cite{2008kweePowe}	2008 Kwee	LKMR binary c	Kernel	Kernel metho	presents a powerful and
\cite{2008liMethods}	2008 Li & Leal	CMC	Burden test	Rare variant	Introduces the combin
\cite{2008liuEstimatic}	2008 Liu	Kernel	Kernel	Kernel machir	The method proposed b
\cite{2008moskvinaM}	2008 Moskva & Sci	Multiple test	Reference me	Multiple testi	This paper by Moskva
\cite{2009madsenGro}	2009 Madsen & Brov	WSS Weighte	Adaptive Bur	Burden tests	Madsen and Browning p
\cite{2009mukhopadH}	2009 Mukhopadhyay	KBAT Associat	Kernel	Kernel machir	The KBAT method prop
\cite{2009panAsympt}	2009 Pan	SSU test The s	Regression	Regression	The SSU method by Pan
\cite{2010hallInnovat}	2010 Hall & Jin	iHC	Enforce spars	Tests that en	Hall and Jin introduce t
\cite{2010hanDataAd}	2010 Han & Pan	aSum	Adaptive bur	Adaptive bur	The aSum method by H
\cite{2010hoffmannCc}	2010 Hoffmann	StepUp	Combined tes	Combined tes	The StepUp method by I
\cite{2010kangVariar}	2010 Kang	EMMAX	Variance com	Variance com	Kang et al.'s EMMAX m
\cite{2010kingEvoluti}	2010 King	EMMPAT Evol	Combined tes	Generalized n	King's EMMPAT metho
\cite{2010liuNovel}	2010 Liu & Leal	KBAC Kernel-b	Kernel	Kernel machir	The method proposed b
\cite{2010logsdonVar}	2010 Logsdon	V-Bay	Bayesian	Bayesian met	Logsdon et al. present a
\cite{2010morrisEvali}	2010 Morris & Zegg	MZ	Review	Review paper	As mentioned previousl
\cite{2010pricePoolec}	2010 Price	VT Pooled ass	Burden test	Burden tests	Price et al. present pool
\cite{2010willerMeta}	2010 Willer	METAL	Meta analysis	Meta-analysis	Presents METAL, a tool
\cite{2010zhouAssoci}	2010 Zhou	LASSO Least a	Enforce spars	Regression, T	Zhou's method uses LAS
\cite{2011basuCompa}	2011 Basu & Pan (Review)	Review	Review	Review paper	This paper by Basu and I
\cite{2011ionita-laza}	2011 Ionita-Laza	RBT Replicatio	Adaptive Bur	Adaptive bur	Ionita-Laza's RBT metho
\cite{2011linGeneral}	2011 Lin & Tang	EREC	Generalized n	The EREC method by Lin	
\cite{2011nealeTestir}	2011 Neale	C-alpha	Variance com	Adaptive bur	Neale's C-alpha test is d
\cite{2011sulOptimal}	2011 Sul	RWAS	Adaptive Bur	Burden tests	Sul's RWAS method is a
\cite{2011wuRareVar}	2011 Wu	SKAT	Variance com	Kernel machir	As previously mentione
\cite{2011yiBayesian}	2011 Yi & Zhi	Bayesian met	Bayesian	Bayesian met	Yi and Zhi introduce a B
\cite{2011zhangFast}	2011 Zhang & Liu	Fast approximation	Regression	Regression	The method proposed b

Variant analysis history

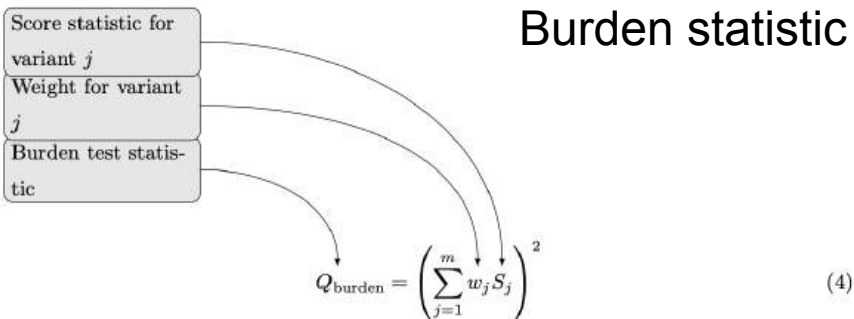
Score statistic



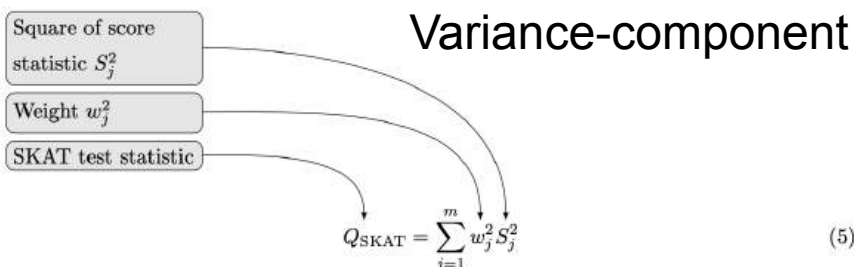
Burden summary



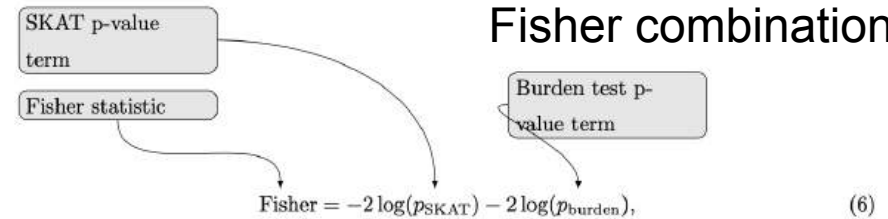
Burden statistic



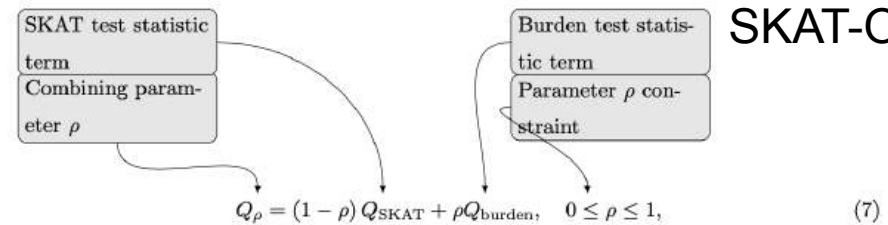
Variance-component



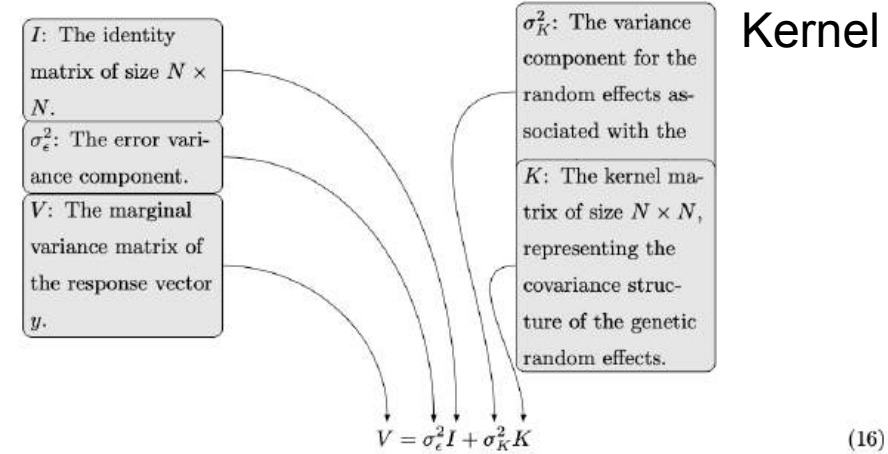
Fisher combination



SKAT-O



Kernel



$$f \sim \text{MVN}(0, \sigma_K^2 K)$$

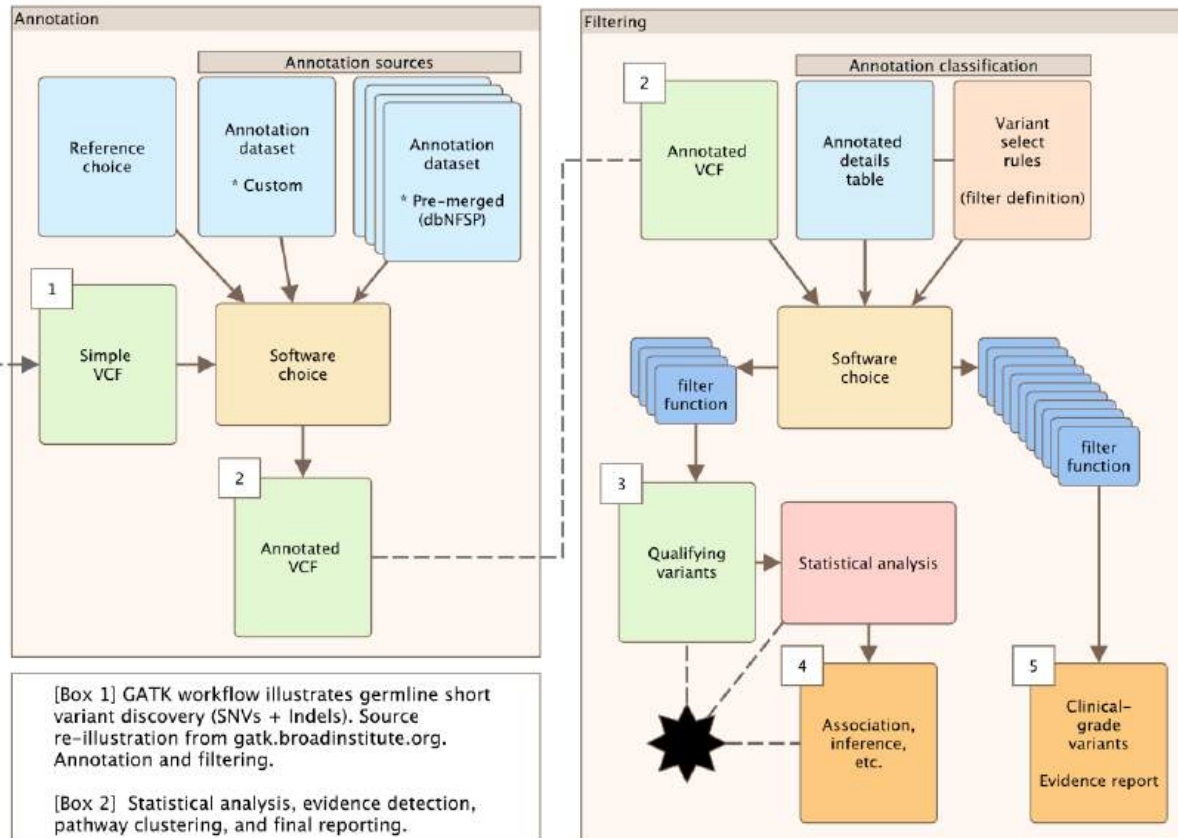
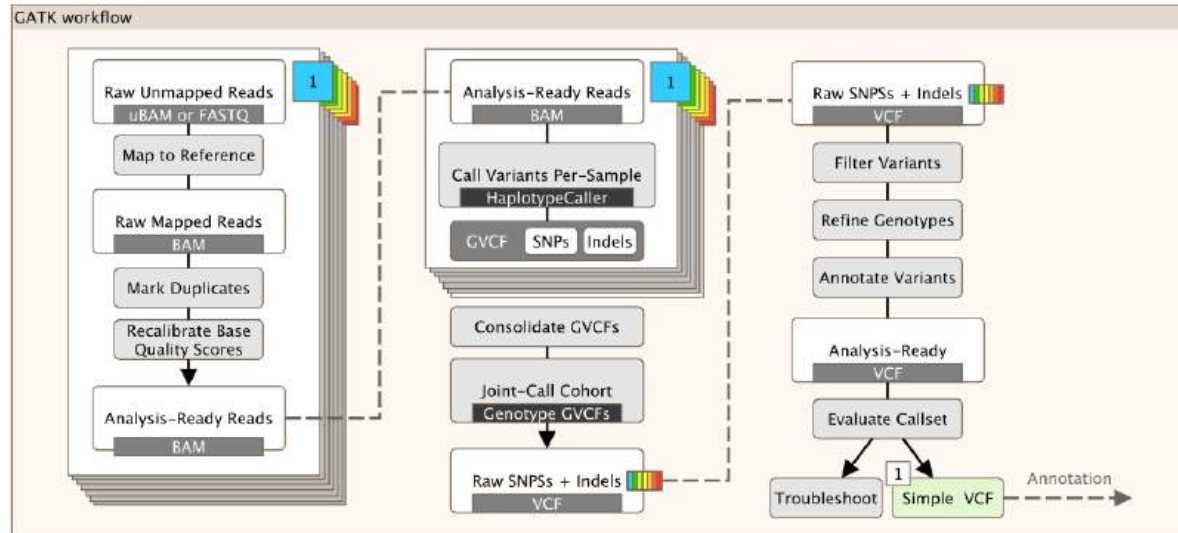
(15)

$$y_i = \alpha_0 + X_i' \alpha + f(G_i) + \epsilon_i$$

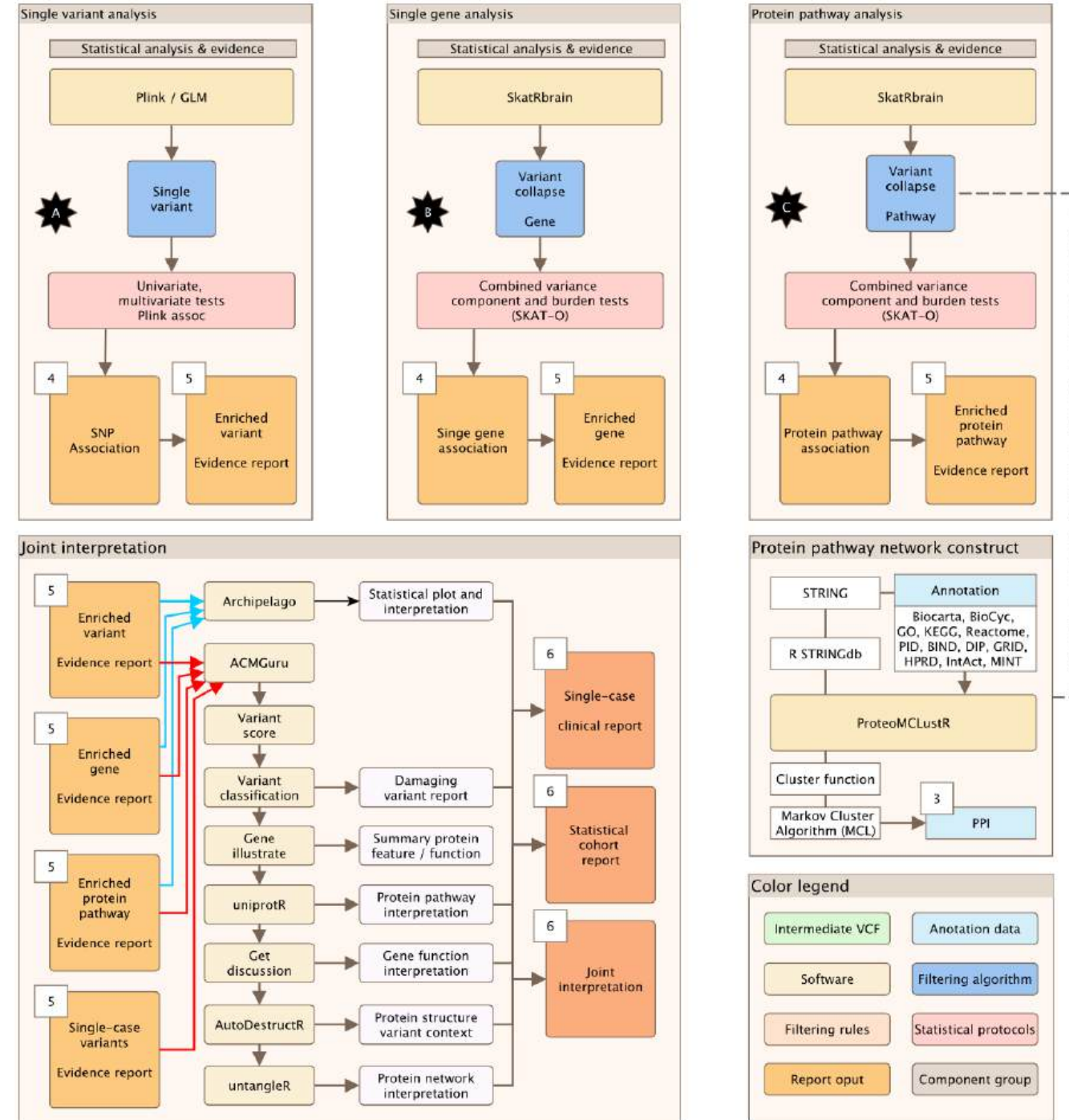
(13)

Modular

Germline variant discovery and association analysis



Statistical analysis and evidence detection



Single case reports

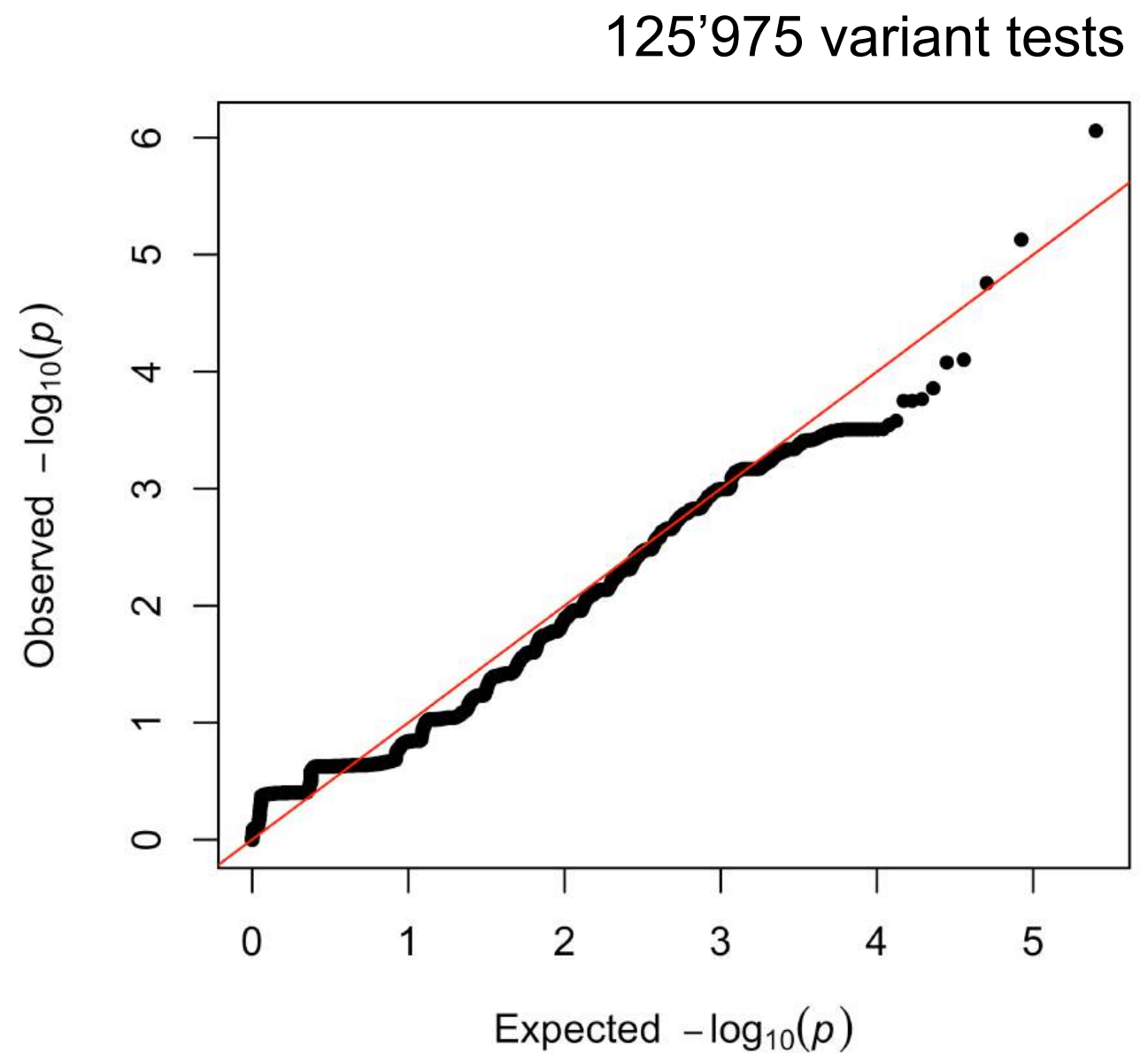
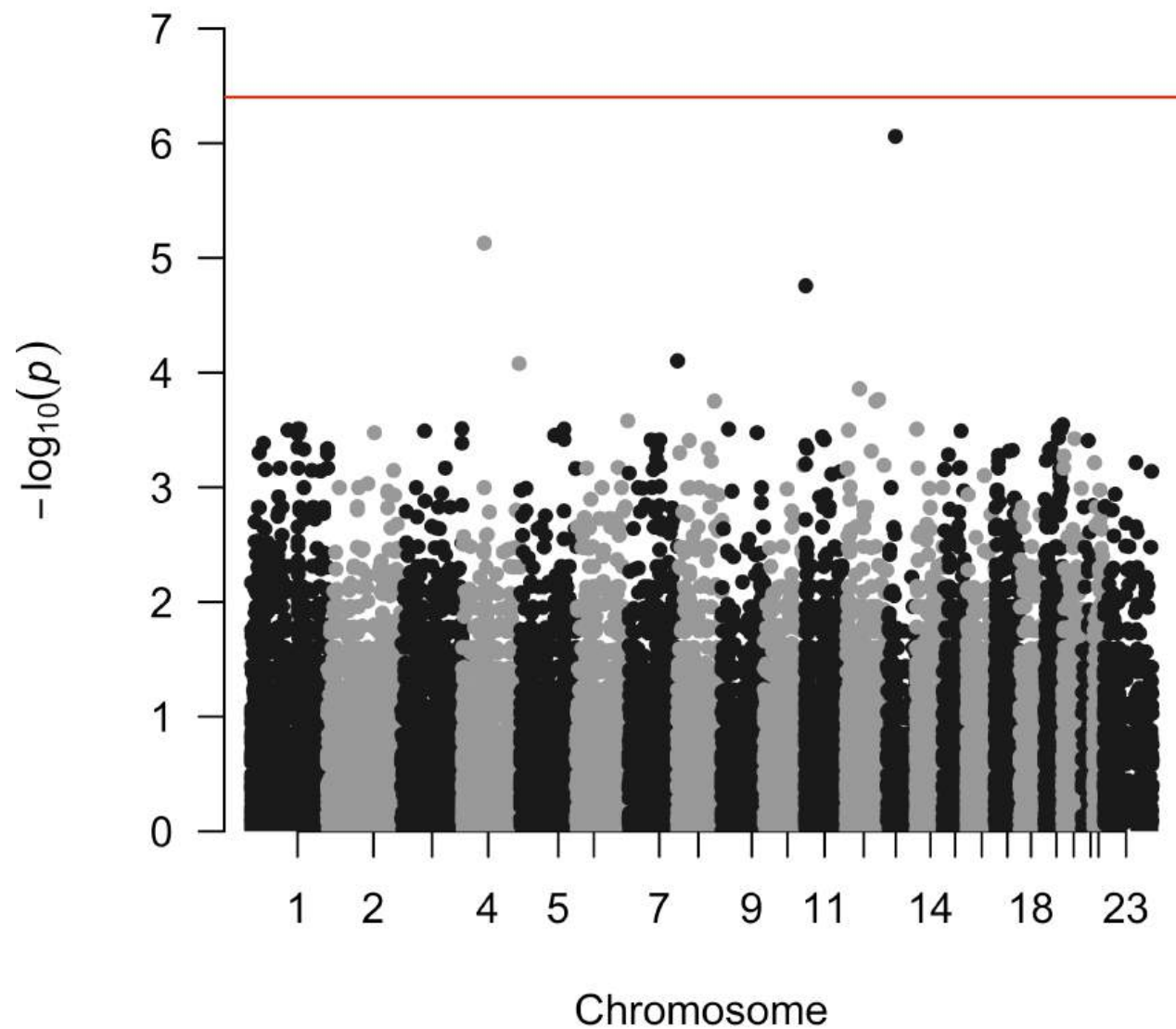
Single-case variants
ACMG/AMP score >6 likely pathogenic
or >10 pathogenic

ACMG score	Consequence	IMPACT	unique_variants
14	stop_gained	HIGH	2
14	splice_acceptor_variant	HIGH	3
14	frameshift_variant	HIGH	6
12	stop_gained	HIGH	1
10	splice_donor_variant	HIGH	1
9	stop_gained	HIGH	3
8	splice_acceptor_variant	HIGH	1
8	stop_gained	HIGH	2
8	stop_lost	HIGH	2
8	splice_donor_variant	HIGH	3
8	frameshift_variant	HIGH	7
7	stop_gained	HIGH	4
7	missense_variant	MODERATE	2
6	splice_acceptor_variant	HIGH	1
6	stop_gained&NMD_transcript_variant	HIGH	1
6	stop_gained	HIGH	2
6	frameshift_variant	HIGH	7
6	missense_variant&splice_region&NMD	MODERATE	1
6	missense_variant	MODERATE	2

Single-case variant genes

Symbol	Consequence
ACD	frameshift_variant
CFHR1	frameshift_variant
CFHR1	frameshift_variant
CFHR1	frameshift_variant
CFHR1	frameshift_variant
CFHR1	splice_acceptor_variant
CFTR	missense_variant
CFTR	splice_donor_variant
IFIH1	frameshift_variant
IFIH1	splice_acceptor_variant
IL17F	frameshift_variant
IL17F	splice_donor_variant
IL17F	splice_donor_variant
NLRC4	stop_gained
NLRP1	frameshift_variant

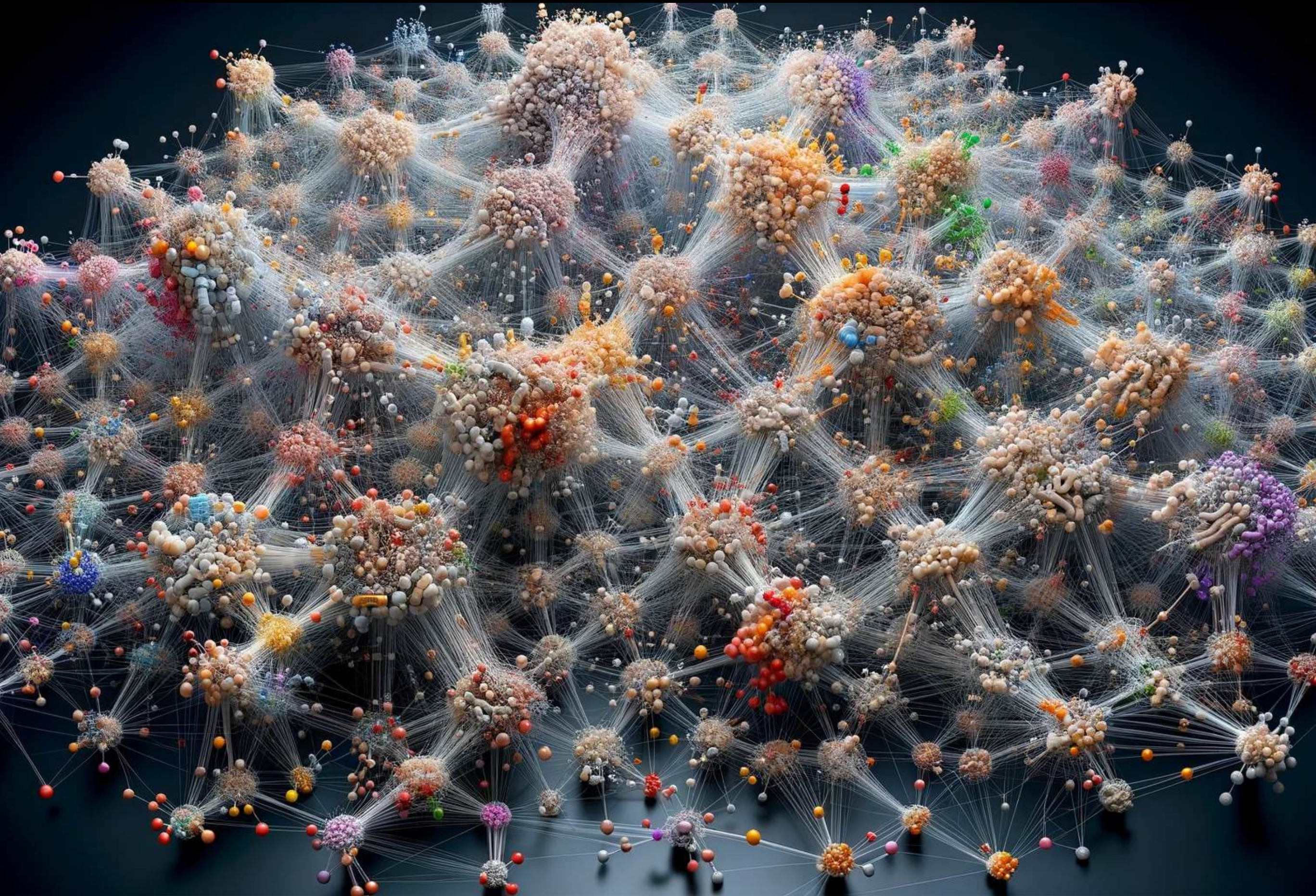
Single variants



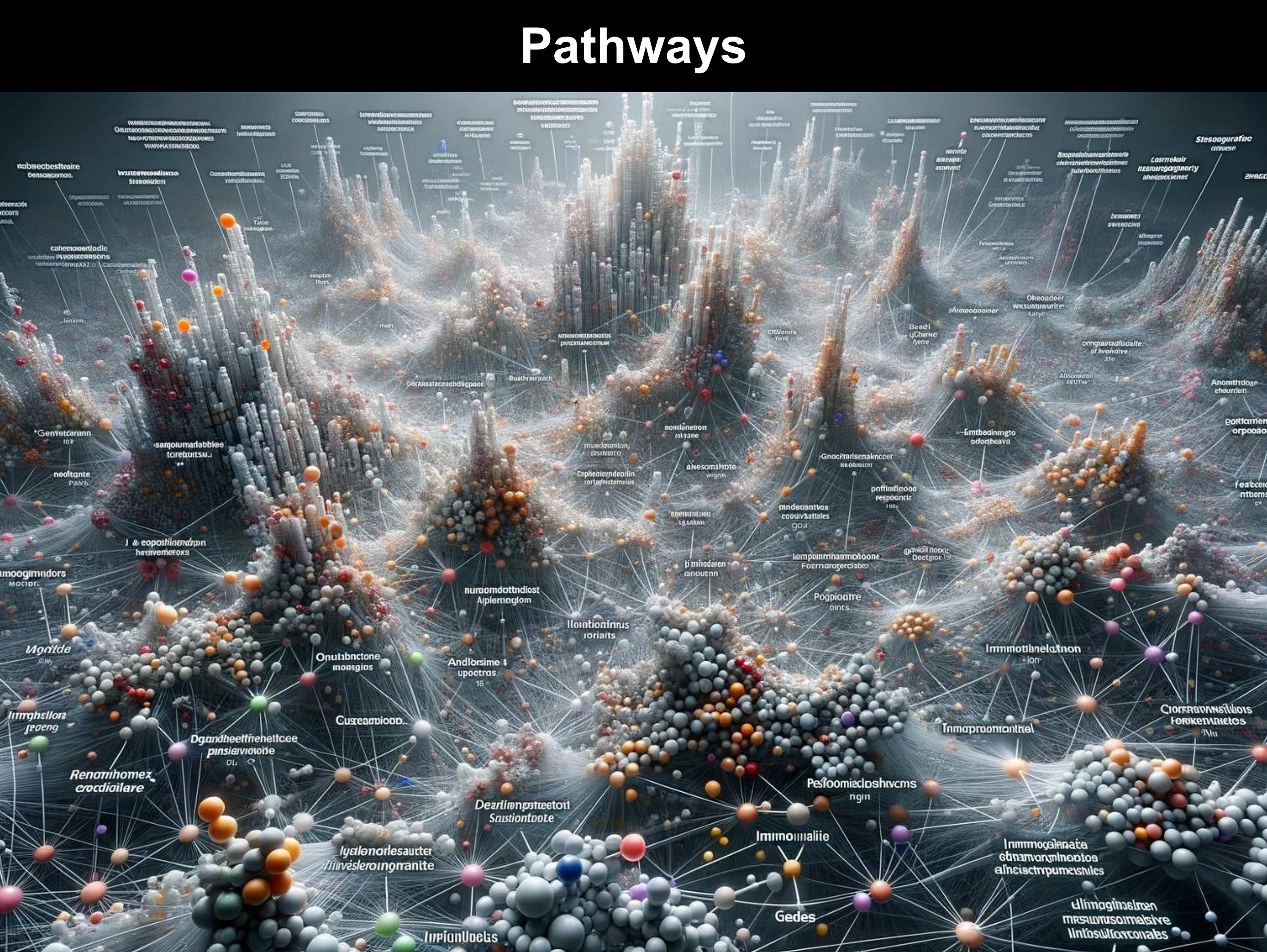
$$\chi^2 = \sum_{i=1}^n \frac{(O_i - E_i)^2}{E_i}$$

$$h_{(\mu_i)} = \alpha_0 \cdot \alpha'_{X_i} + \beta'_{G_i}$$

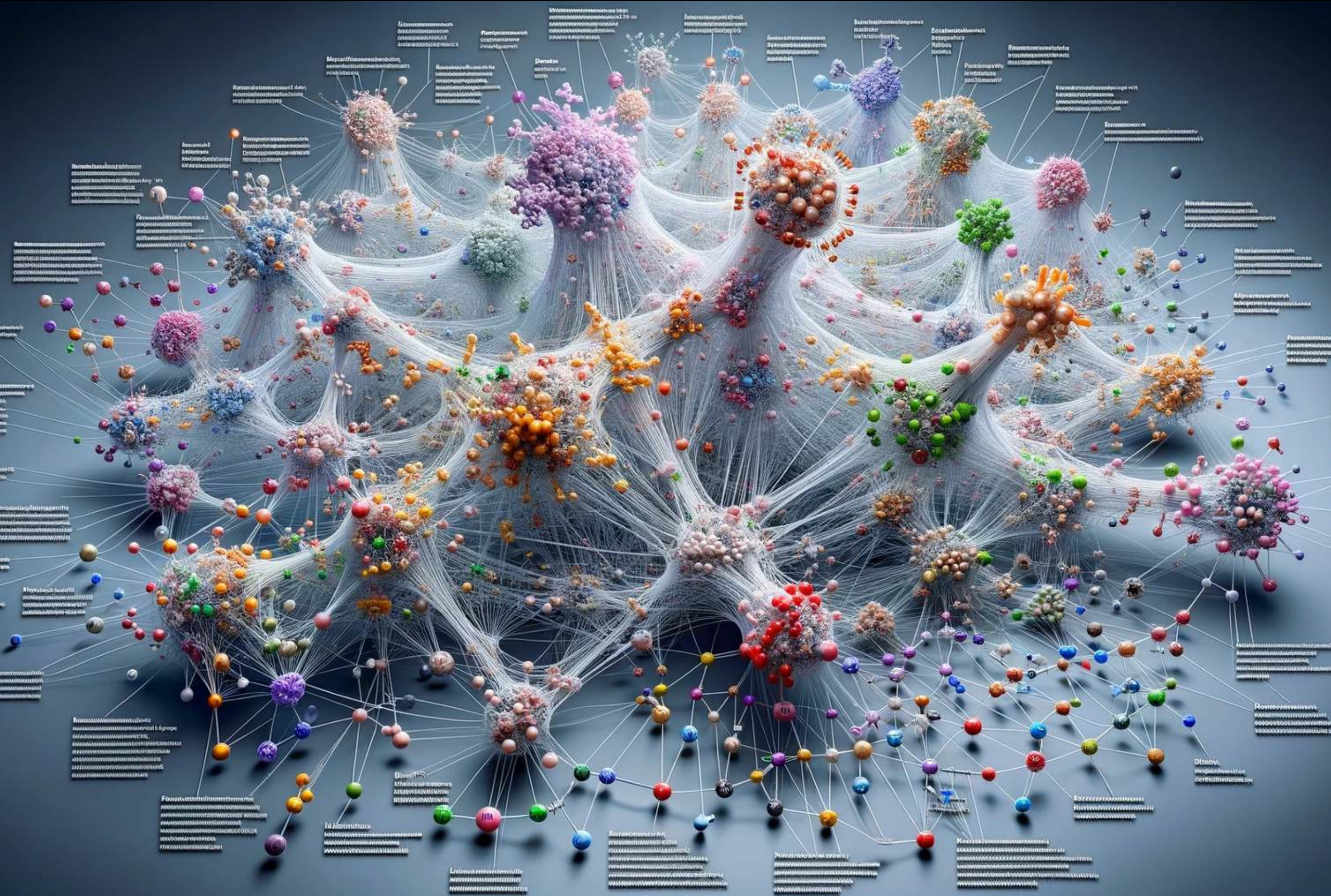
Pathways



Pathways



Pathways



ProteoMCLustR

Input:

$N_i, i = 1, \dots, n :$

Nodes (genes) in the STRING database

$E_{ij}, i, j = 1, \dots, n :$

Edges (interactions) between nodes in STRING database

S : Score threshold for edges

I : Iteration limit

$L_{\min}, L_{\max}$: Size limits for clusters

e, r : Expansion & inflation parameters for MCL algorithm

Algorithm:

1. Preprocess $(N_i, E_{ij}, S) \rightarrow (N'_i, E'_{ij})$
2. ChooseInflation $(N'_i, E'_{ij}, L_{\min}, L_{\max}) \rightarrow \text{inflation}$
3. RunMCL $(N'_i, E'_{ij}, I, L_{\min}, L_{\max}, \text{inflation}, e, r)$

3.1. Initialize $M_{ij}^{(0)} = \frac{E'_{ij}}{\sum_{k=1}^n E'_{ik}}$

3.2. Iterate until convergence:

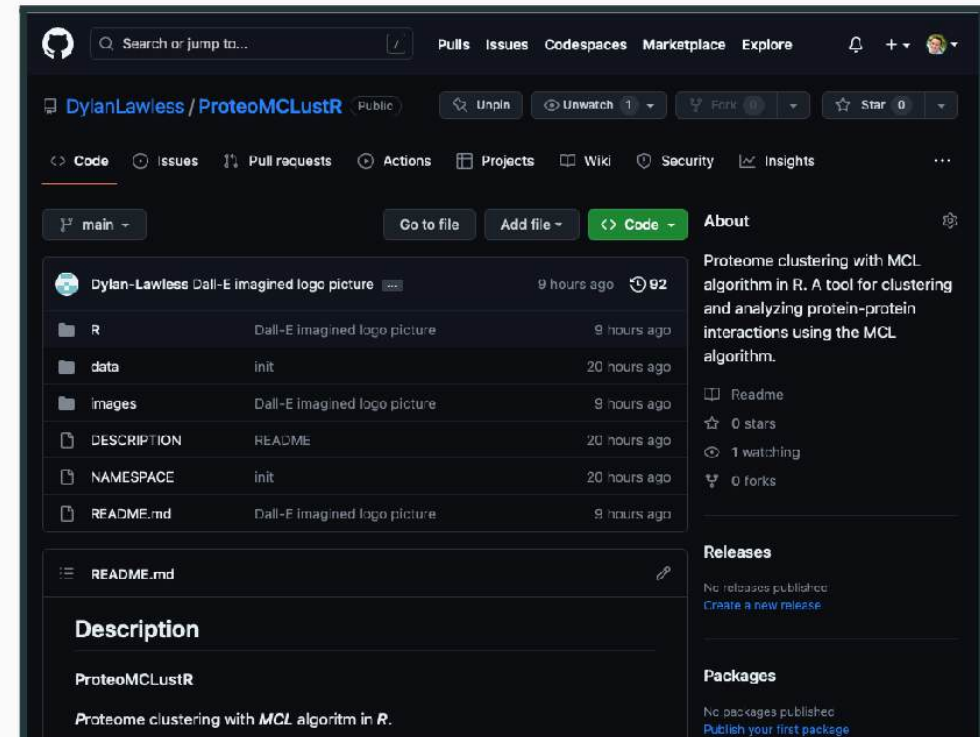
3.2.1. Expansion $M^{(k)} = (M^{(k-1)})^e$

3.2.2. Inflation $M_{ij}^{(k)} = \frac{(M_{ij}^{(k-1)})^r}{\sum_{k=1}^n (M_{ik}^{(k-1)})^r}$

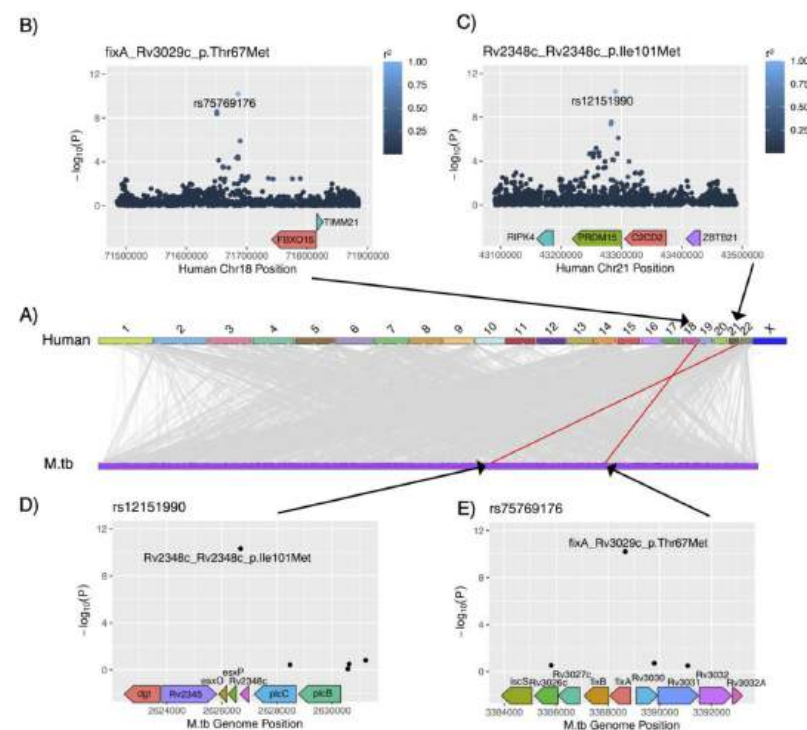
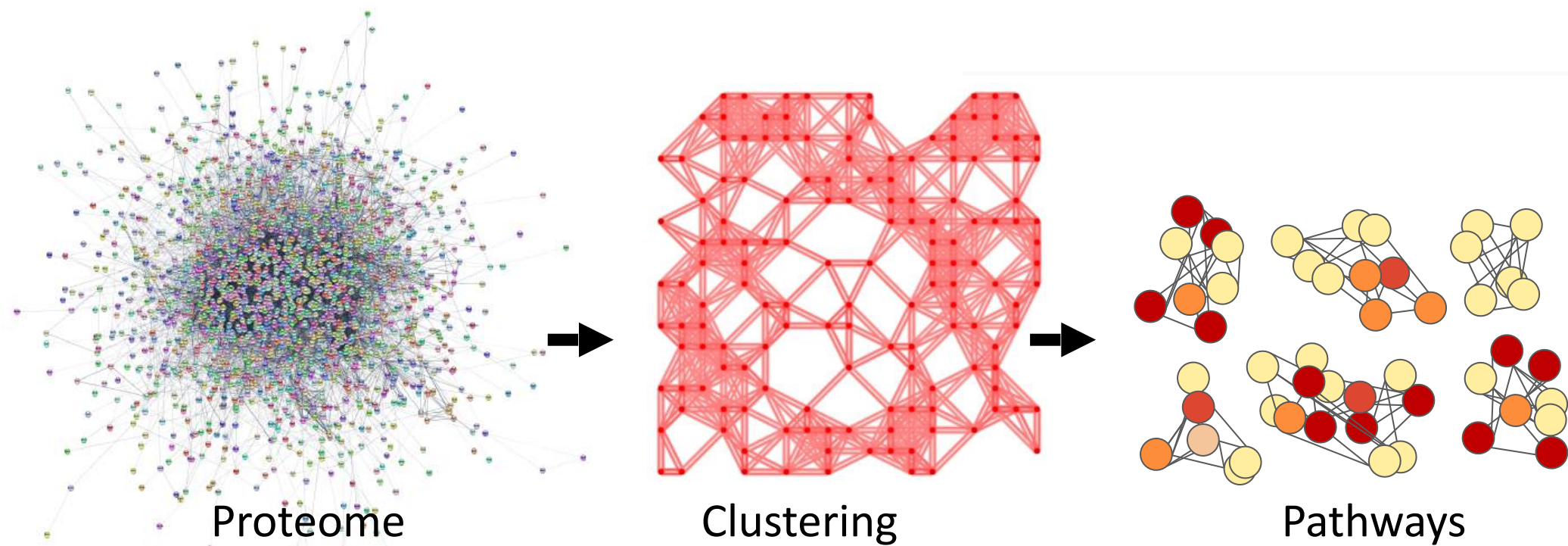
3.3. Extract clusters from converged matrix $M^{(\text{final})}$

Output:

Clusters: Set of optimized node (gene) clusters



ProteoMCLustR

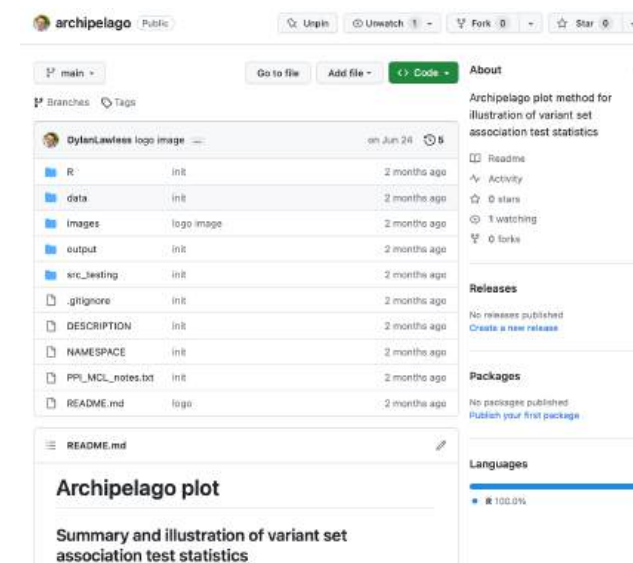


Genome-to-genome analysis reveals associations between human and mycobacterial genetic variation in tuberculosis patients from Tanzania

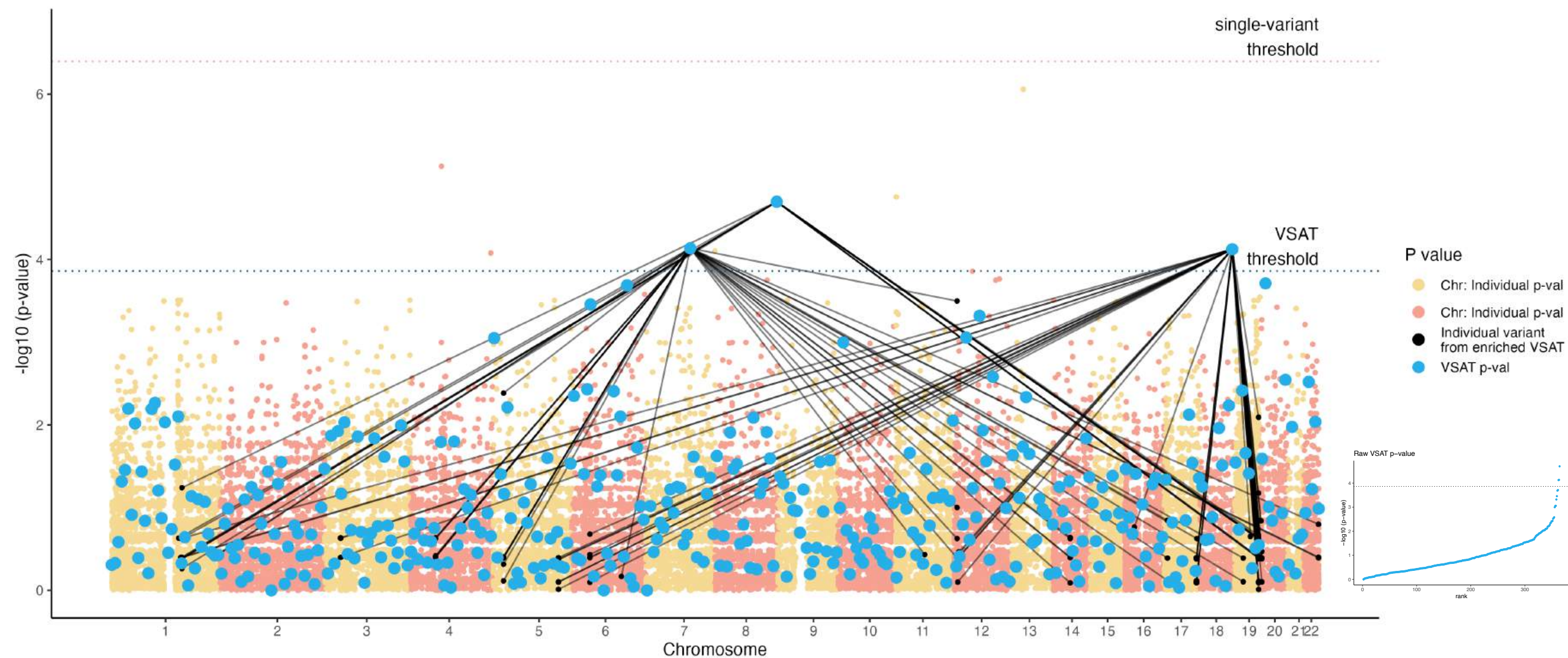
Zhi Ming Xu, Michaela Zwyrer, Daniela Brites, Hellen Hiza, Mohamed Sasamalo, Miriam Reinhard, Anna Doetsch, Sonia Borrell, Olivier Naret, Sina Rüeger, Dylan Lawless, Faima Ishaka, Hosiana Temba, Thomas Maroa, Rastard Naftari, Christian Beisel, Jerry Hella, Klaus Reither, Damien Portevin, Sebastien Gagneux, Jacques Fellay
doi: <https://doi.org/10.1101/2023.05.11.23289848>



Archipelago



Archipelago

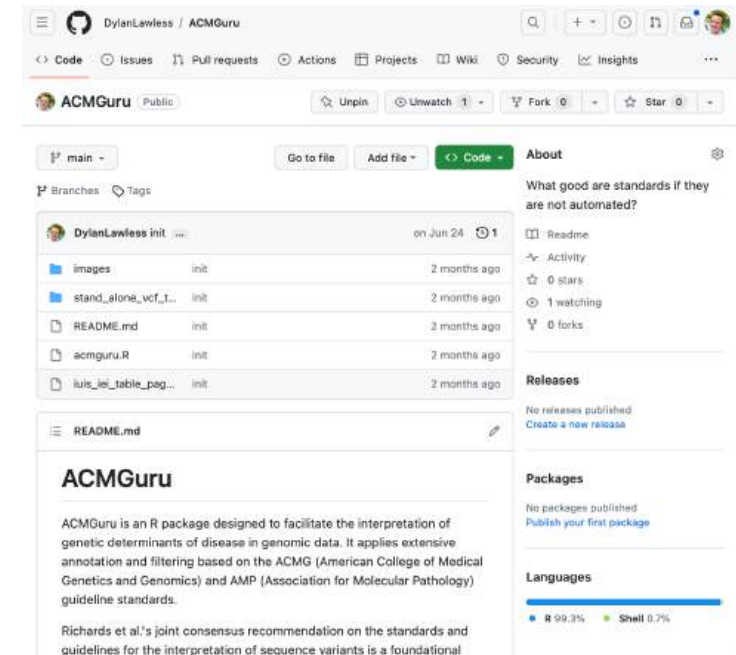


Pre-print. Archipelago plot method for illustration of variant set association test statistics. 2023 Lawless and Fellay.

Evidence

ACMGuru

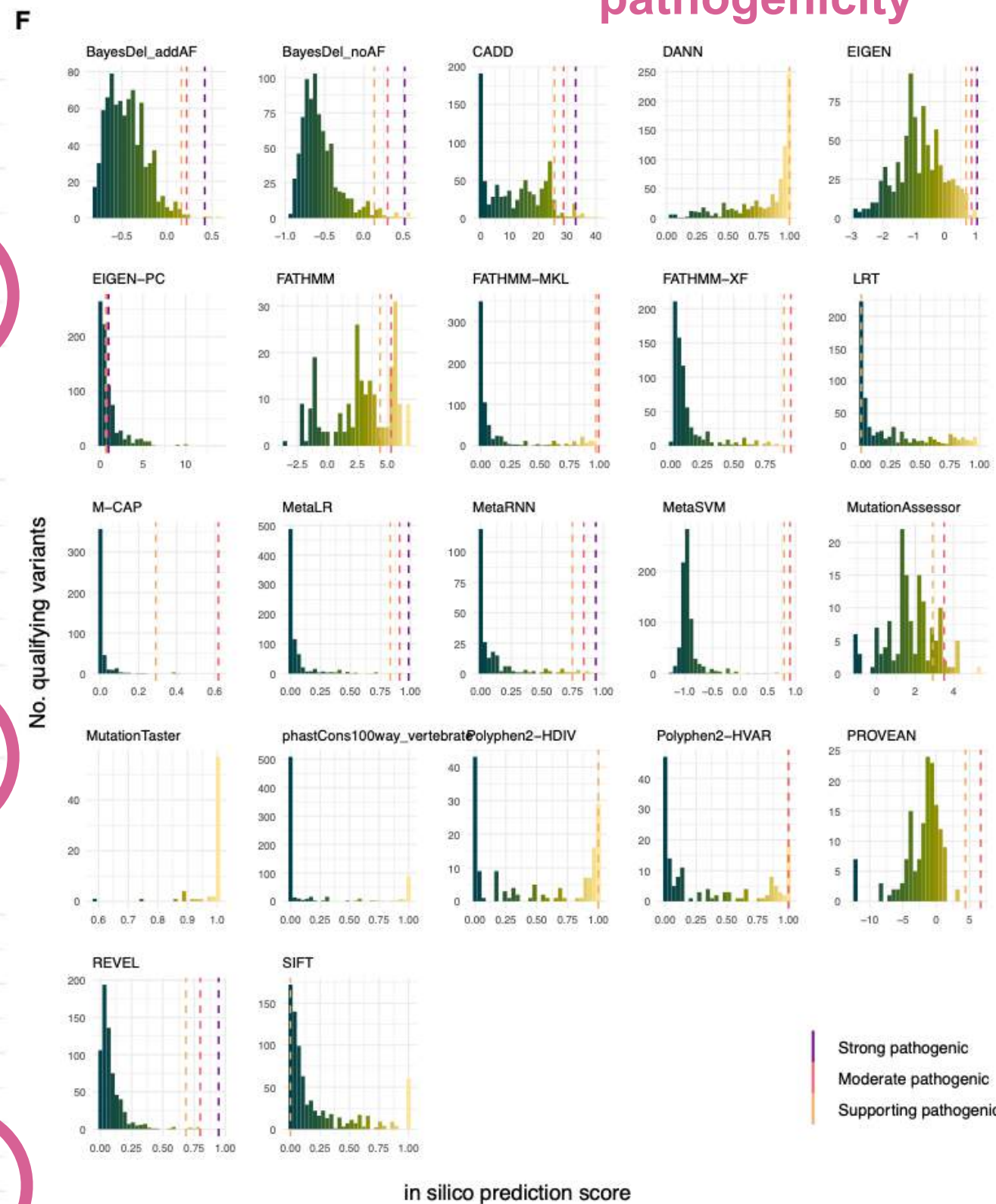
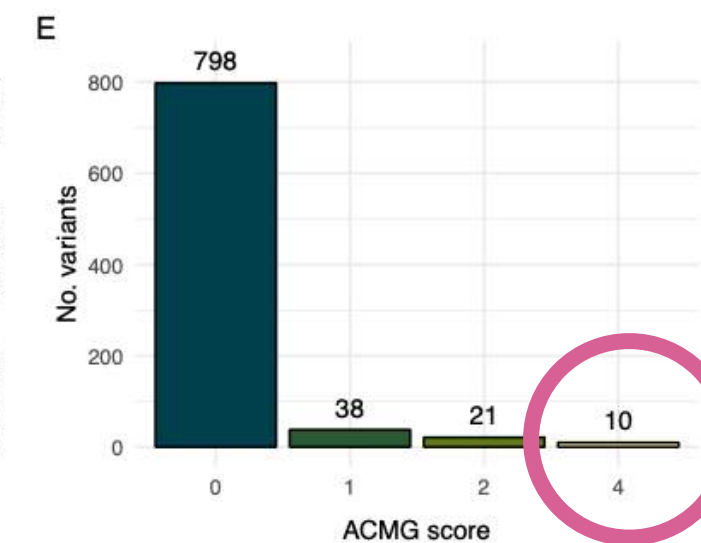
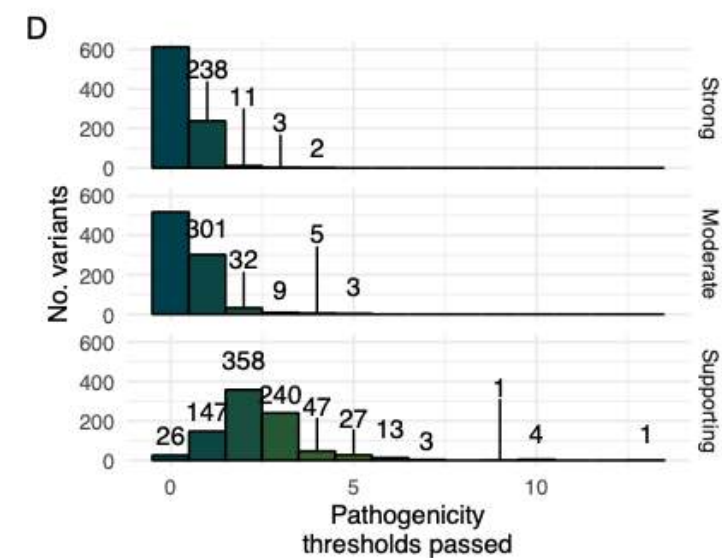
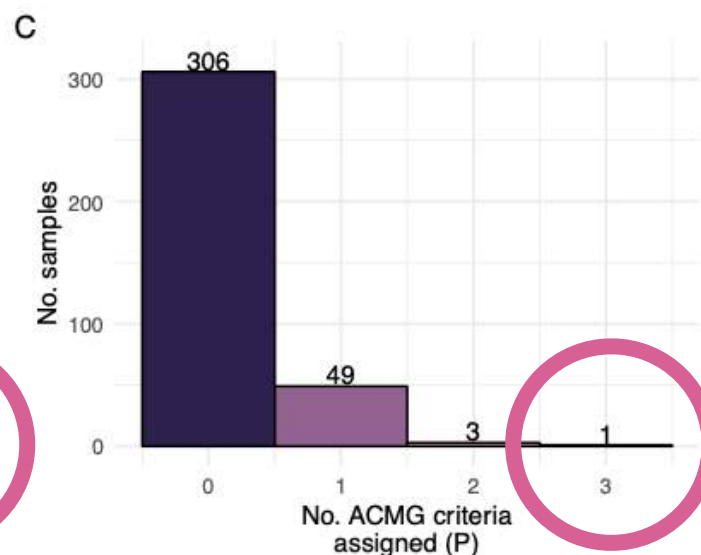
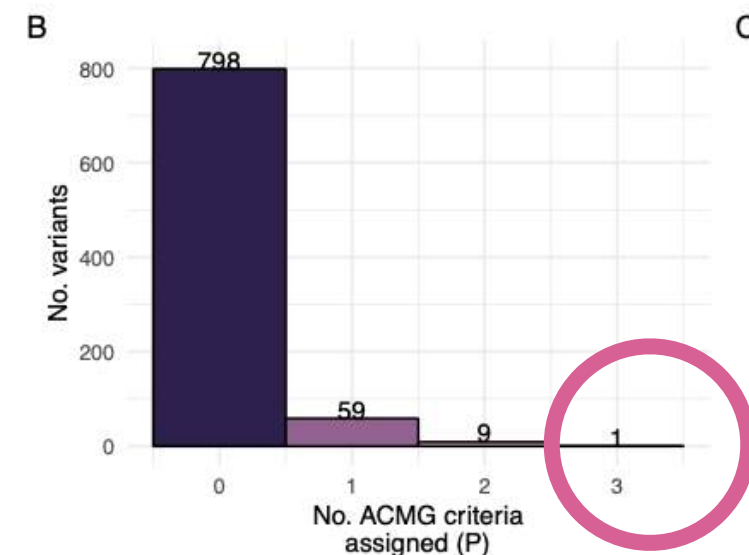
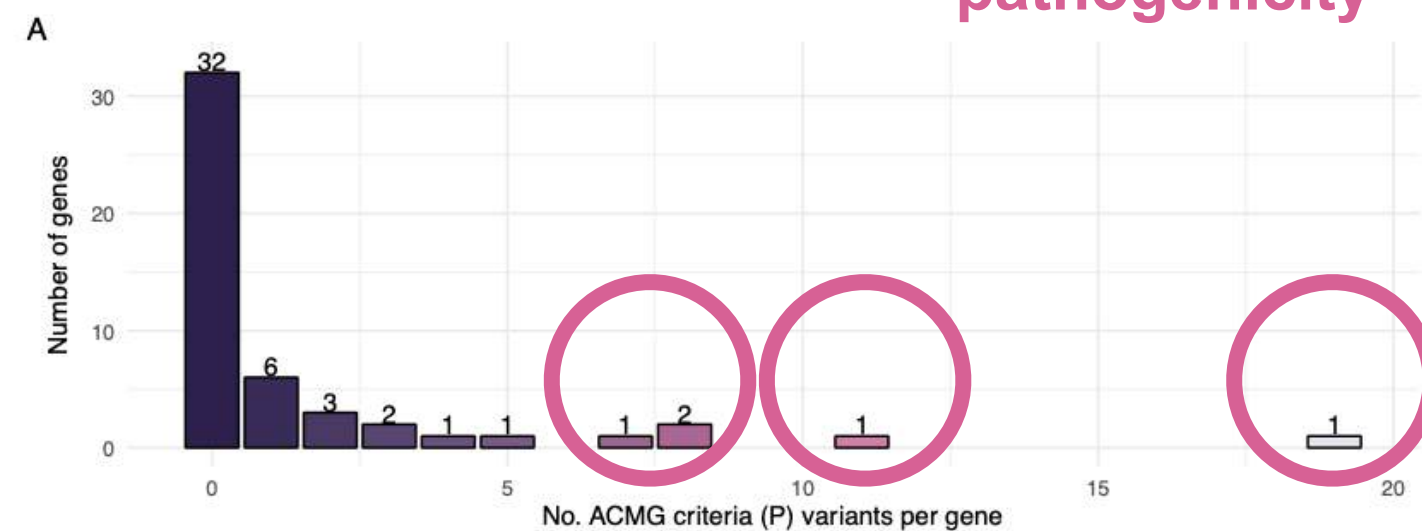
- Joint consensus recommendation on the standards and guidelines for the interpretation of sequence variants (Richards 2015).
- Standards and guidelines for the interpretation and reporting of sequence variants in cancer (Li 2017).
- Technical standards for the interpretation and reporting of constitutional copy-number variants (Riggs 2020).
- Effective variant filtering and expected candidate variant yield in studies of rare human disease (Pedersen 2021).
- ...



Evidence: scoring

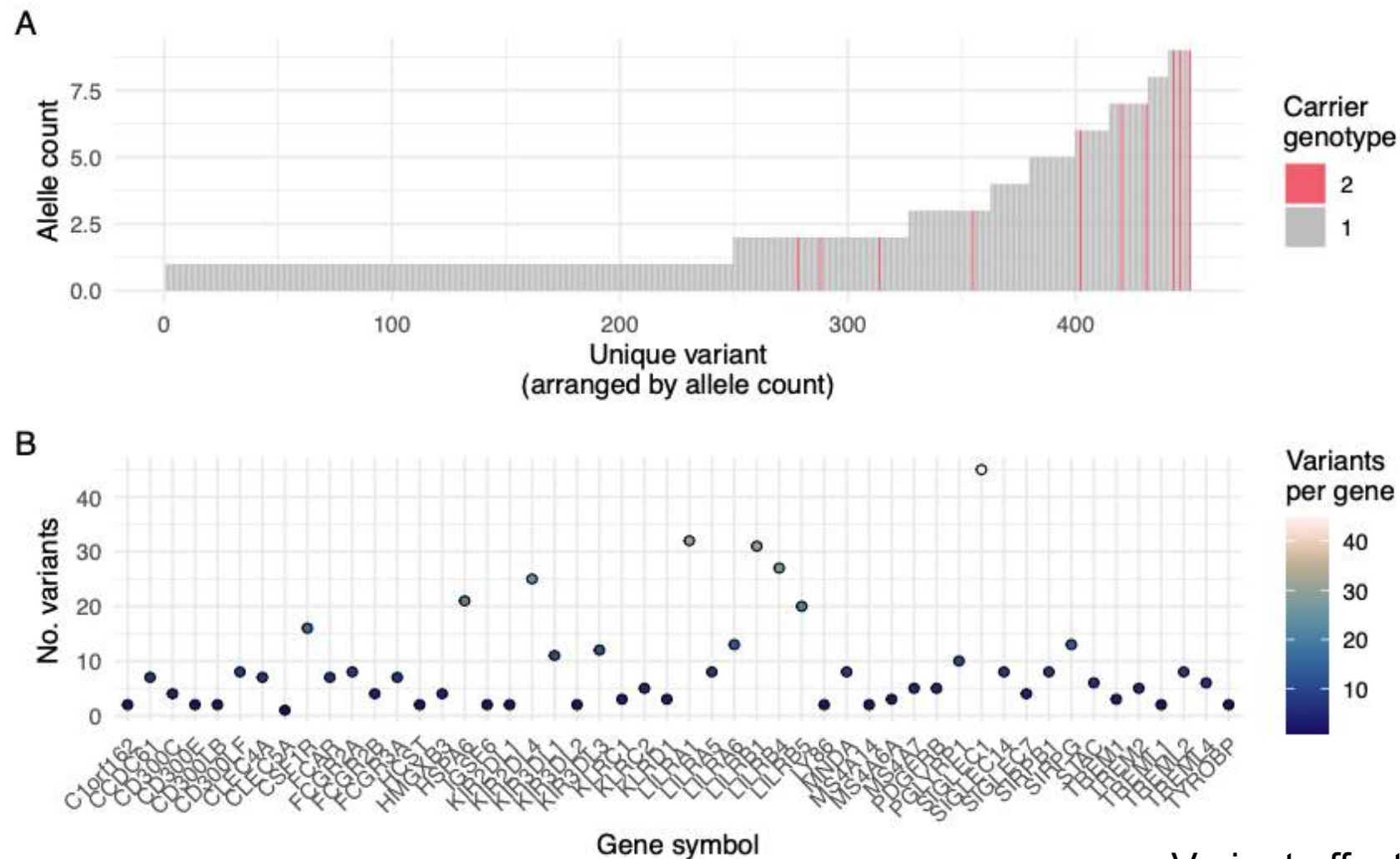
Evident
pathogenicity

Predicted
pathogenicity



Evidence: technical

Distribution within enriched pathway



Enriched pathway variants,
ACMG/AMP score >6, likely pathogenic

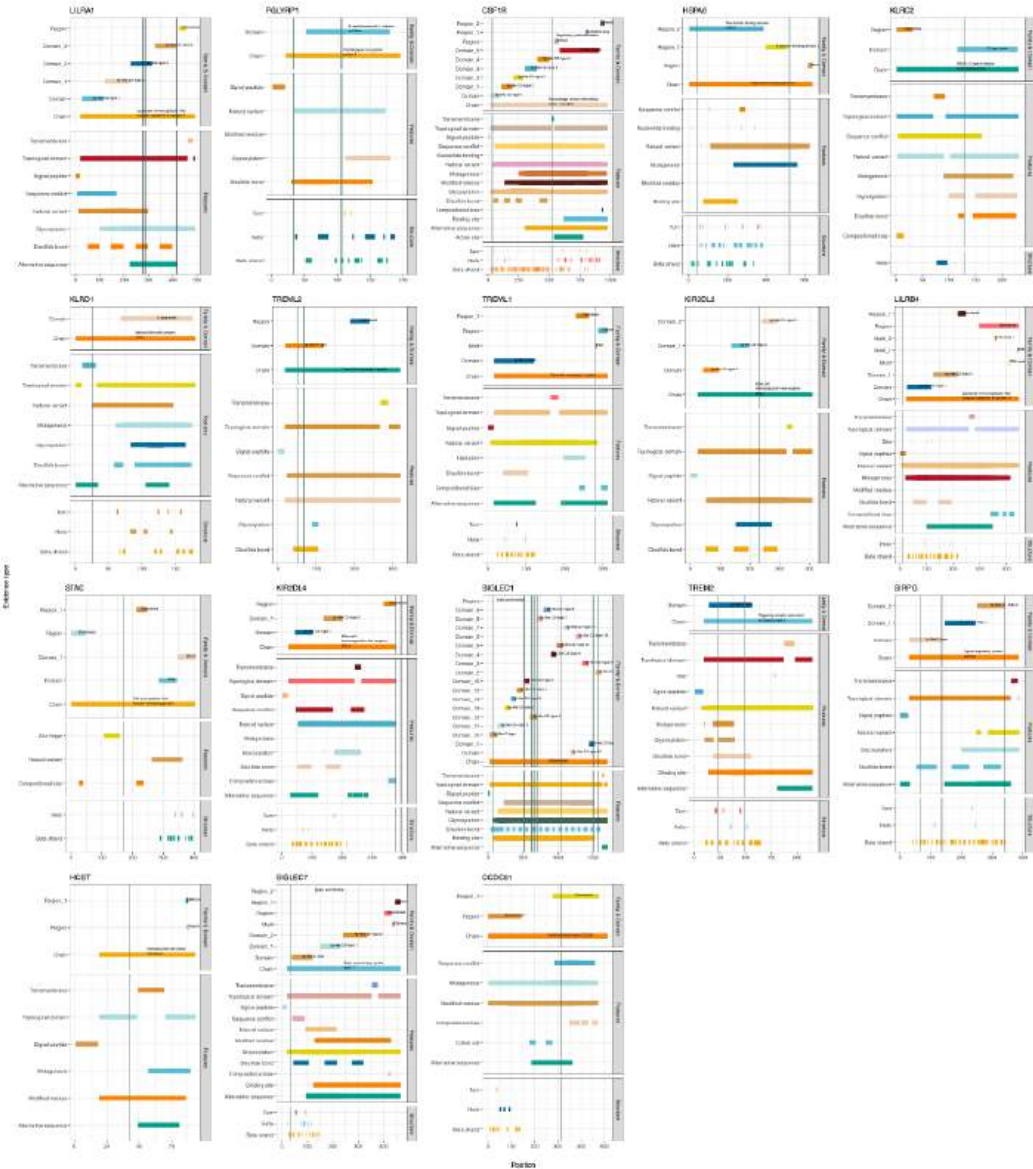
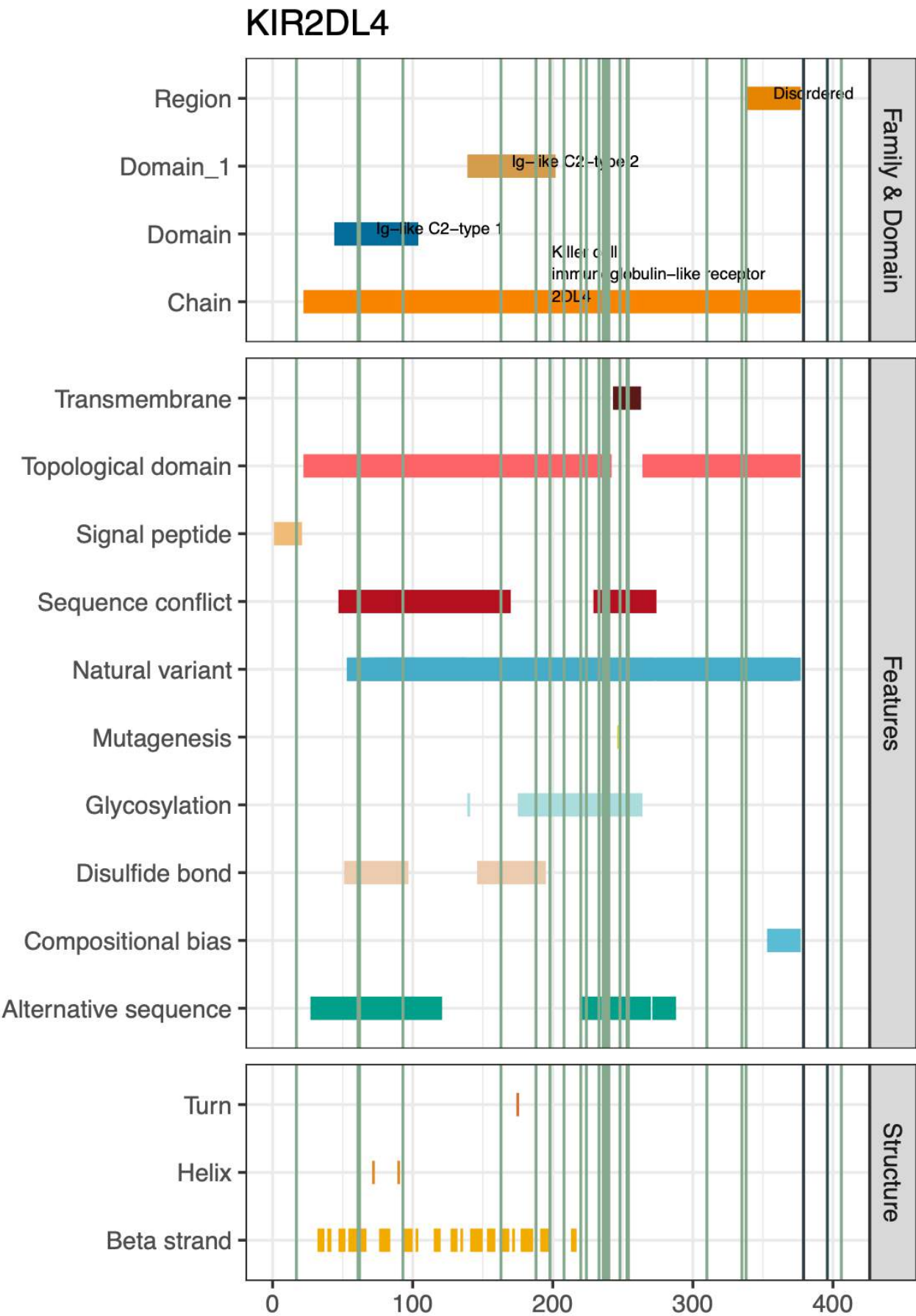
SYMBOL	HGVSp	Consequence	sample.id	genotype
KIR2DL4	ENSP00000379580.3:p.Ala379Val	missense_variant	GE076	1
KIR2DL4	ENSP00000379580.3:p.Asp310LysfsTer5	frameshift_variant	GE076	1
LILRA1	ENSP00000251372.3:p.Glu416Asp	missense_variant	P501	1
LILRA1	ENSP00000251372.3:p.Glu416Gly	missense_variant	P501	1
LILRA1	ENSP00000251372.3:p.Leu277Pro	missense_variant	ZH114	1
LILRA1	ENSP00000251372.3:p.Ser414Pro	missense_variant	P501	1
LILRA1	NA	splice_acceptor_variant	P501	1
LILRA1	NA	splice_acceptor_variant	ZH114	1
SIRPG	ENSP00000305529.3:p.Arg342Cys	missense_variant	LAU176	1
SIRPG	ENSP00000305529.3:p.Pro136GlnfsTer29	frameshift_variant	LAU176	1

Variant effect summaries

Consequence	IMPACT	genotype	unique_variants
frameshift_variant&splice_region_variant	HIGH	1	1
start_lost	HIGH	1	1
stop_gained&splice_region_variant	HIGH	1	1
stop_lost	HIGH	1	1
splice_donor_variant	HIGH	1	3
splice_acceptor_variant	HIGH	1	4
frameshift_variant	HIGH	1	8
stop_gained	HIGH	1	17
inframe_deletion&splice_region_variant	MODERATE	1	1
inframe_deletion	MODERATE	1	2
missense_variant&splice_region_variant	MODERATE	1	9
missense_variant	MODERATE	2	9
missense_variant	MODERATE	1	305

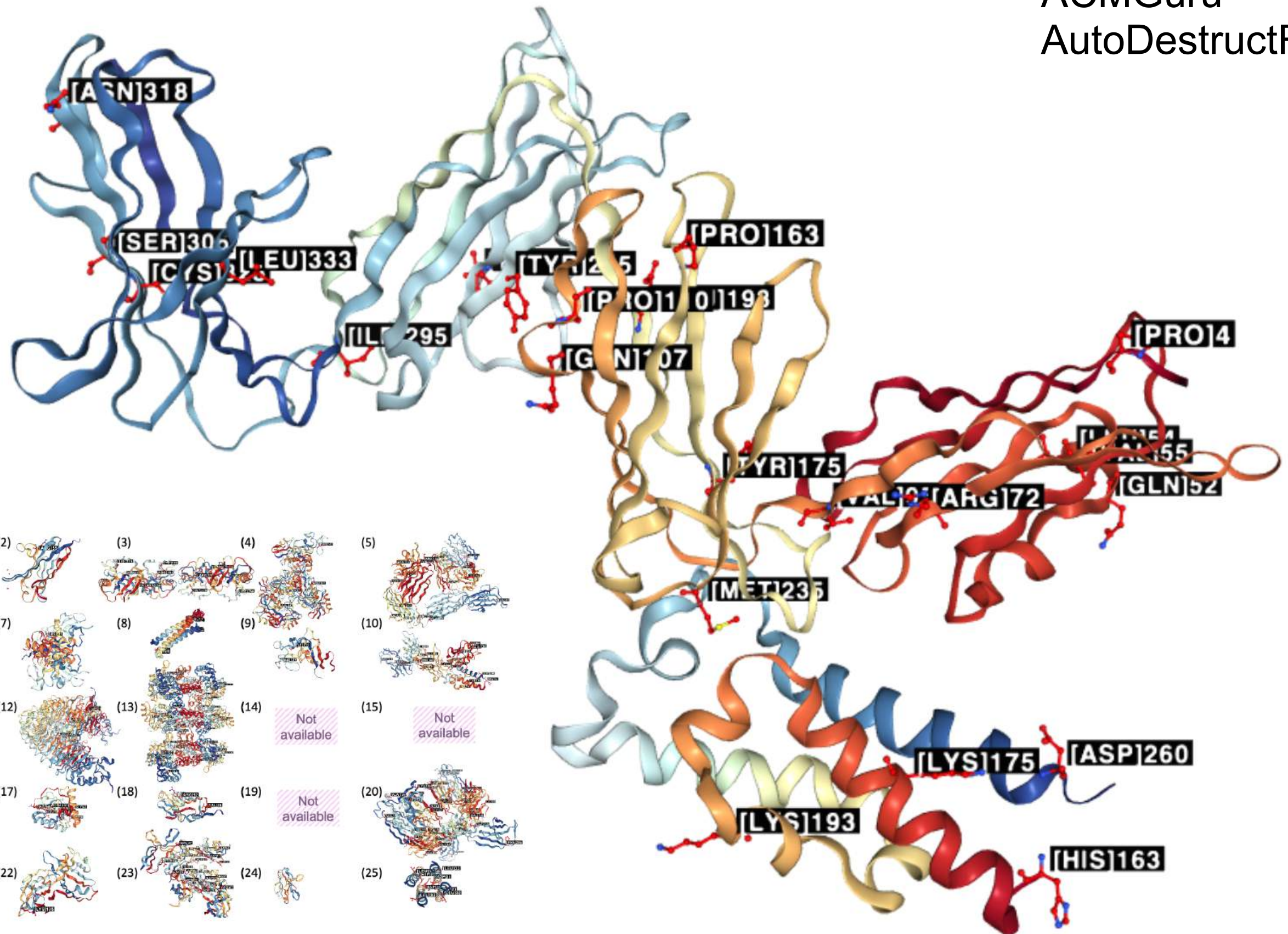
Evidence: protein function

ACMGuru
gene-illustrate



Evidence: protein structural

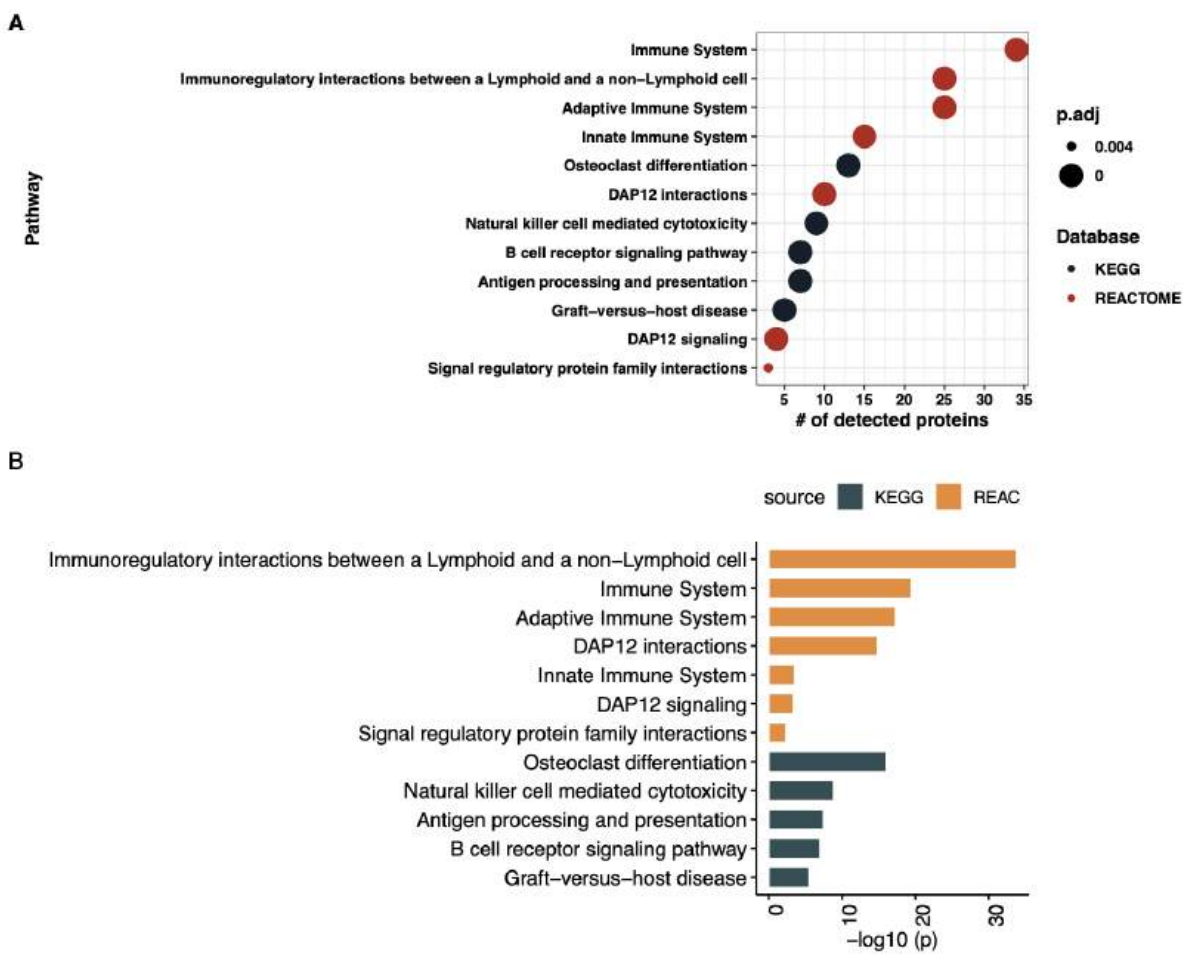
ACMGuru
AutoDestructR



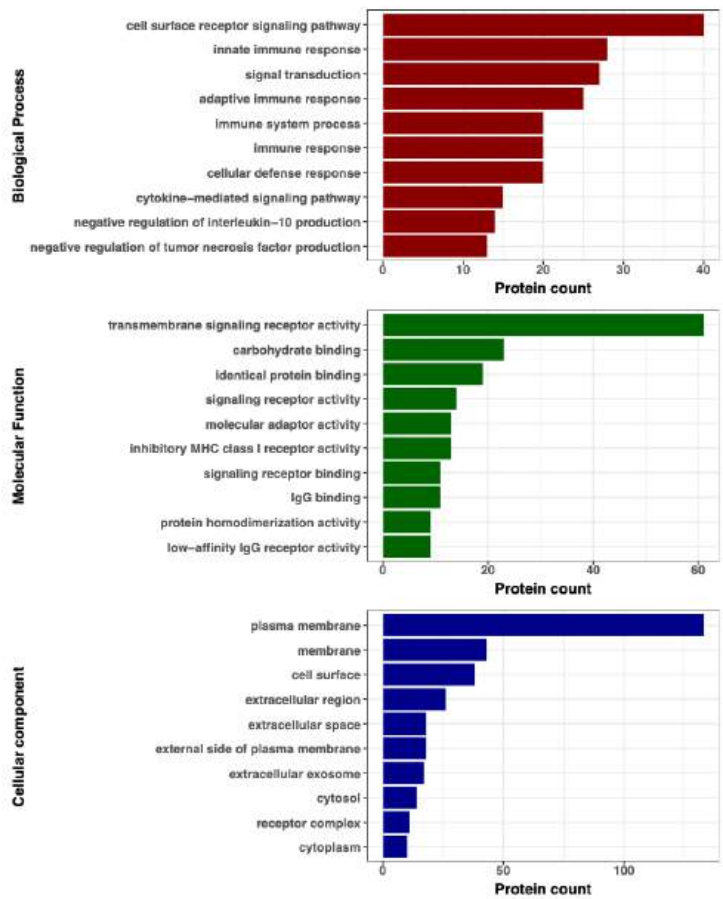
Evidence: pathway function

ACMGuru
uniprotR

Kegg and Reactome pathway interpretation



Gene Ontology (GO) pathway interpretation



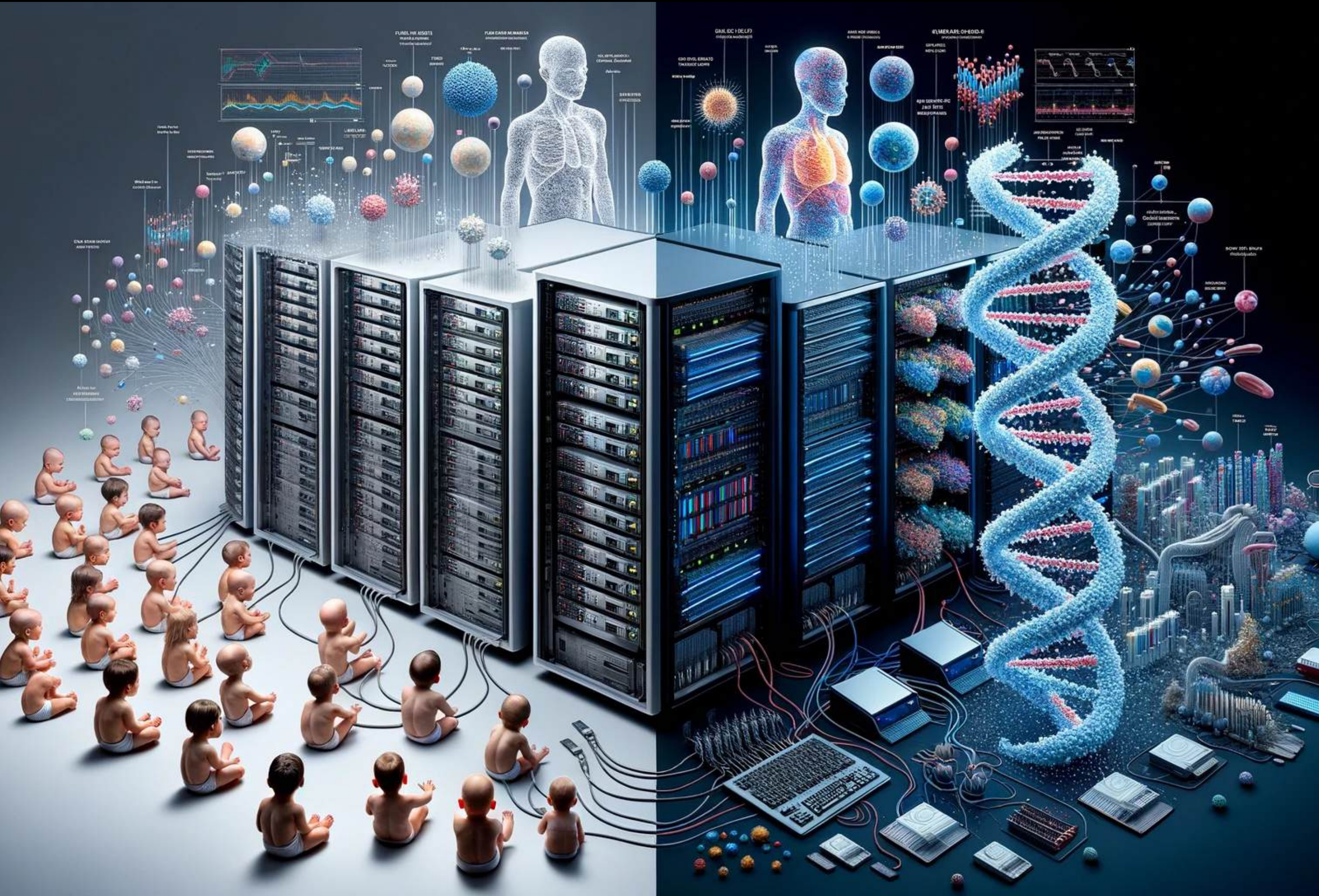
Goterm	Count	%
cell surface receptor signaling pathway	40	22.22%
innate immune response	28	15.56%
signal transduction	27	15.00%
adaptive immune response	25	13.89%
cellular defense response	20	11.11%
immune response	20	11.11%
immune system process	20	11.11%
cytokine-mediated signaling pathway	15	8.33%
negative regulation of interleukin-10 production	14	7.78%
negative regulation of tumor necrosis factor production	13	7.22%

Goterm	Count	%
transmembrane signaling receptor activity	61	33.89%
carbohydrate binding	23	12.78%
identical protein binding	19	10.56%
signaling receptor activity	14	7.78%
inhibitory MHC class I receptor activity	13	7.22%
molecular adaptor activity	13	7.22%
IgG binding	11	6.11%
signaling receptor binding	11	6.11%
low-affinity IgG receptor activity	9	5.00%
protein homodimerization activity	9	5.00%

Goterm	Count	%
plasma membrane	133	73.89%
membrane	43	23.89%
cell surface	38	21.11%
extracellular region	26	14.44%
external side of plasma membrane	18	10.00%
extracellular space	18	10.00%
extracellular exosome	17	9.44%
cytosol	14	7.78%
receptor complex	11	6.11%
cytoplasm	10	5.56%

Now

Scale and test



Scale and test



Scale and test

Cluster Specs

- 419 compute nodes
- 2 Intel(R) per node (36 cores each)
- 3 TB SSD disk
- 512 GB RAM/node min

Max Requirements

- Time: 24 hours
- 24 CPUs/node; 80GB per CPU
- 1000 job array
- Occasionally 500GB RAM

Runtimes for 1000 sample

- GATK pipeline: < 8 hours
- Clustering: < 10 hours (parallel to GATK)
- Statistical analysis: < 2 hours



A face behind the research



Thanks

Intensivmedizin & Neonatologie

Wir betreuen Kinder und Jugendliche, deren Gesundheitszustand lebensbedrohlich gefährdet ist oder werden könnte. Dazu setzen wir auf modernste Geräte, innovative Behandlungsmethoden und speziell ausgebildete sowie erfahrene Fachpersonen.



Organizing committee

Reto Huber (Chair)
Matthias Baumgartner
Micaela Belendez
Barbara Brotschi
Maya Horst
Karin Kucian
Nicole Meili

Poster jury

Patrick Meyer (Chair)
Martin Baumgartner
Sophie Böttcher
Barbara Brotschi
Suzanne Dhaini
Sean Froese
Raphael Morscher
Jana Pachlopnik
Paolo Paioni

Michelle Seiler
Ana Stücklin
Martin Theiler
Huub van Hedel
Flavia Wehrle
Stephan Wüest

Oral jury

Matthias Baumgartner
Martin Baumgartner
Barbara Brotschi
Sean Froese
Johannes Häberle
Reto Huber
Maya Horst
Uchenna Kennedy
Karin Kucian
Nicole Meili
Ola Sabet

Prof Luregn Schlapbach

Prof Matthias Baumgartner

Prof Jacques Felley



SwissPedHealth Lighthouse project

