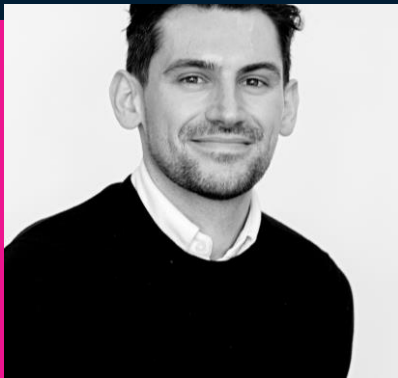




MEDINFO23

8 – 12 JULY 2023 | SYDNEY, AUSTRALIA



@Mitchob

Polygenic Epistatic Phenotype Simulation (PEPS)

Mitchell O'Brien

Postdoctoral Fellow
Transformational Bioinformatics
CSIRO



AICH



Genomic information transforms health care

- Diagnostics
- Treatment
- Screening
- Health risk prediction
- Molecular Tracking



Global Alliance for Genomics and Health (GA4GH) predicts **half a billion people** will be sequenced by 2030



Security



Time constraints

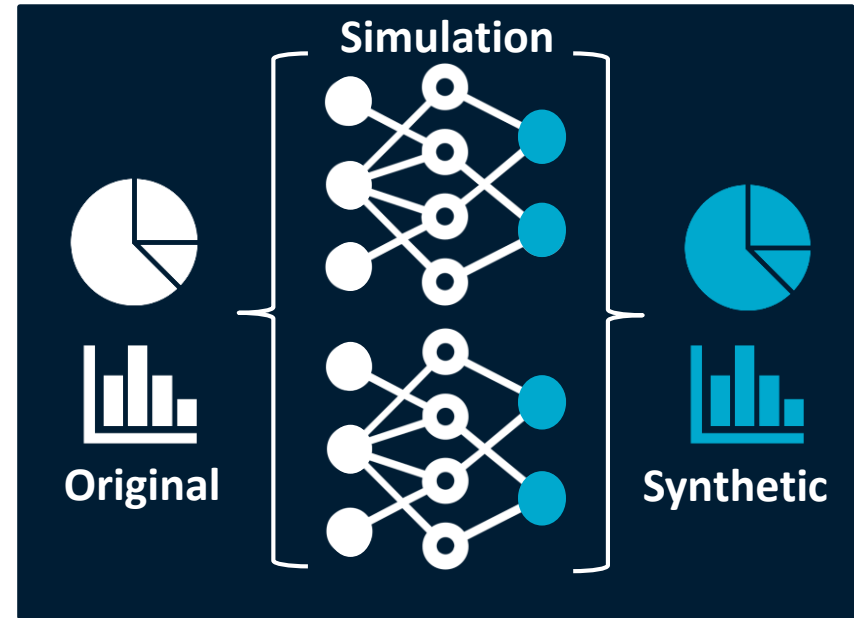


Expensive

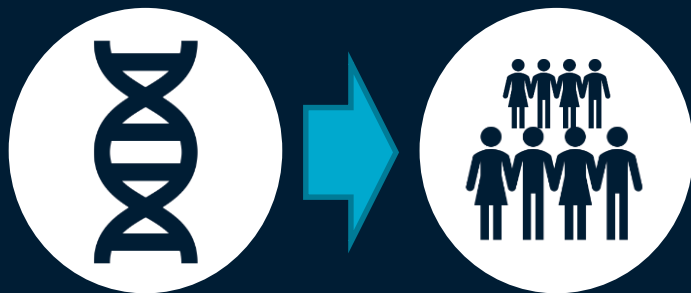
Barriers to technical innovation

Solution – Synthetic data

- “Data Generated by a computer simulation”
- Has the same utility as the original data sources
- Mimics sensitive data without holding sensitive information – privacy preserving
- In genomics - emulate different postulated mechanisms of inheritance

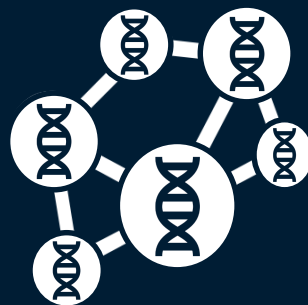


Polygenic Epistatic Phenotype Simulator (PEPS)

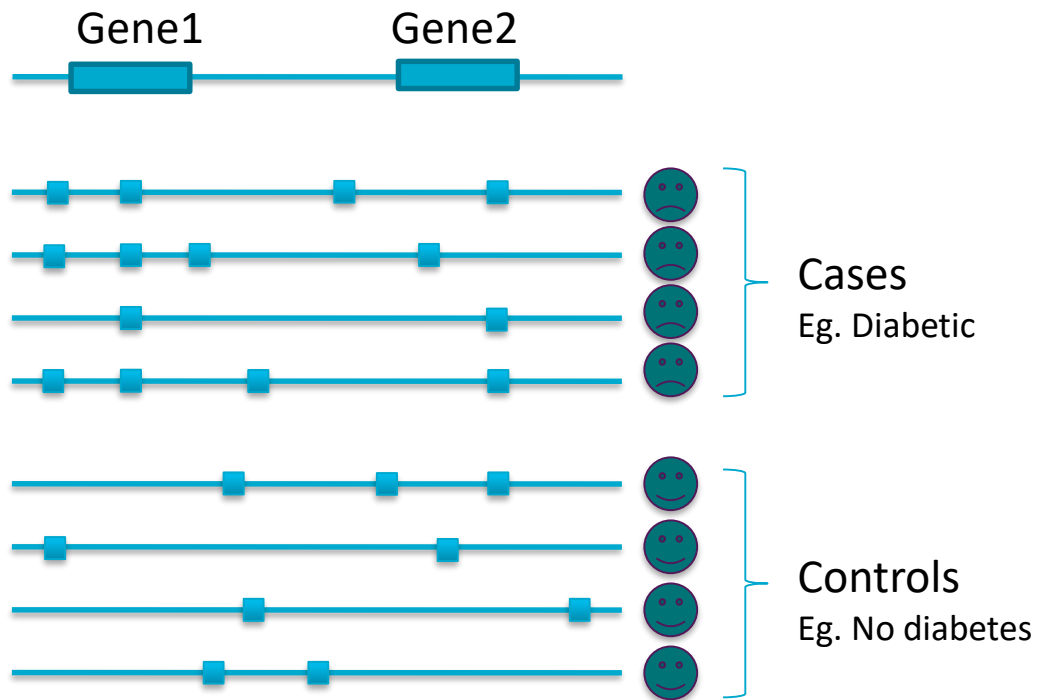


Produces synthetic phenotypes using real genotype data

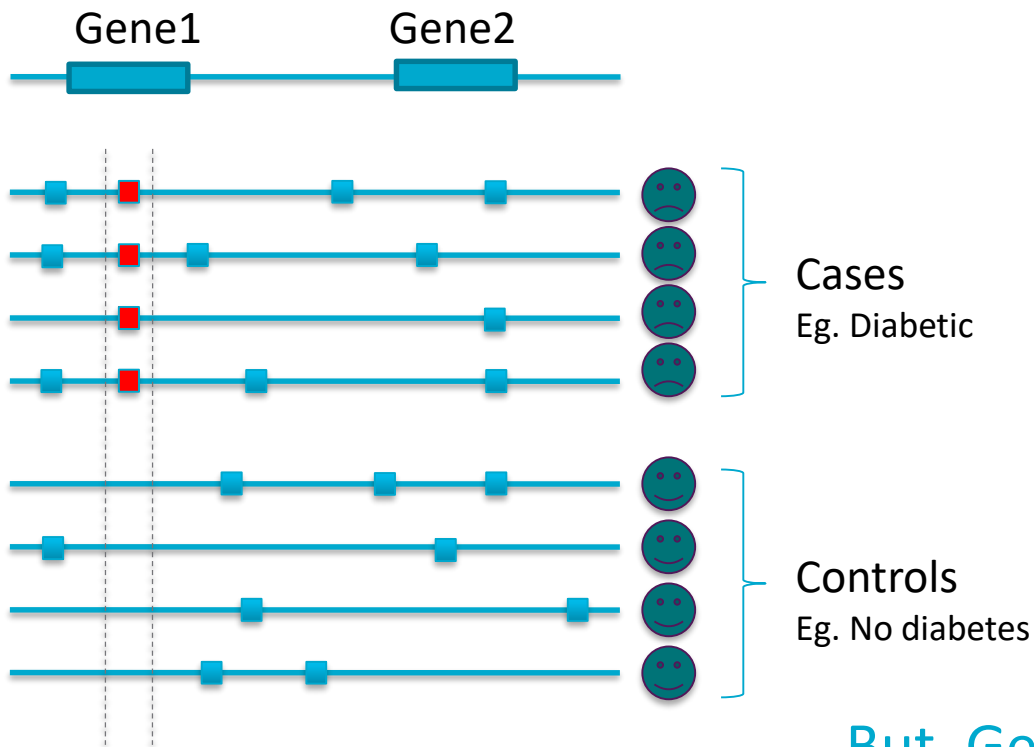
Able to simulate increasingly high order interactions



Epistasis

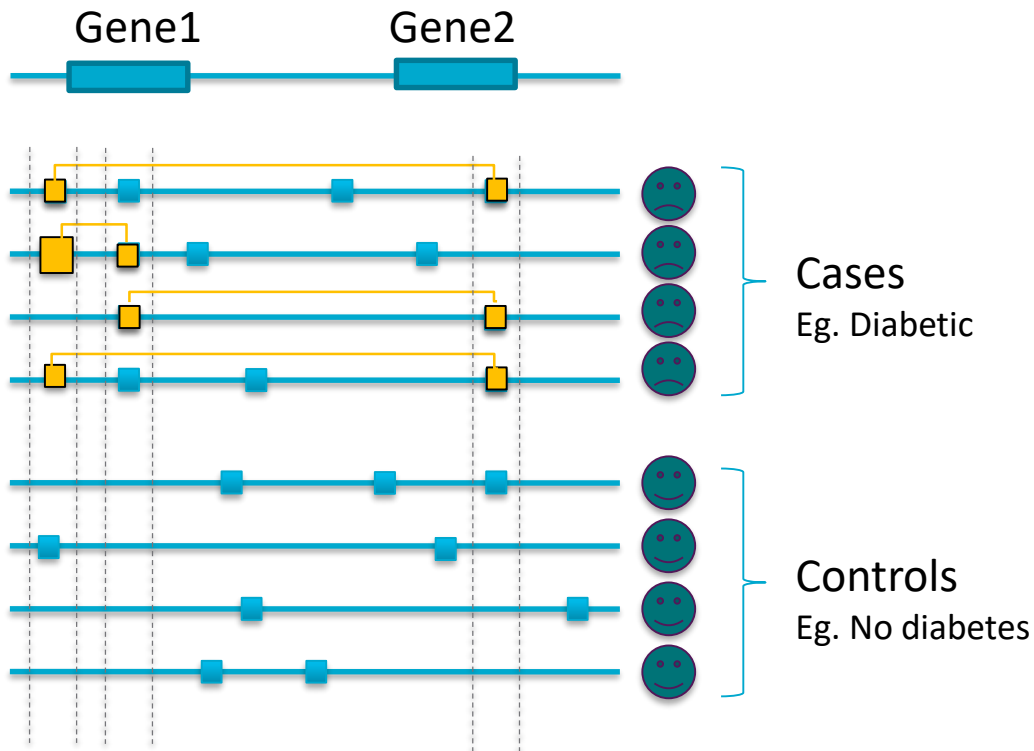


Epistasis



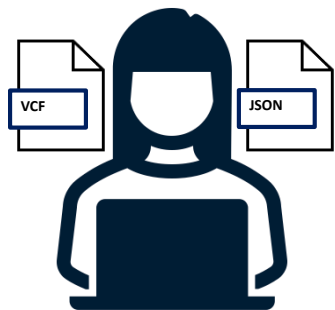
But, Genetics is not simple

Epistasis



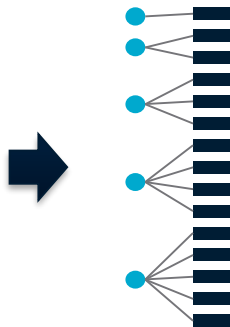
PEPS is **not limited** by the number of interactions a user wants to simulate and produces phenotypes where **single variants and multiple interacting variants** contribute to the final dataset.

PEPS algorithm



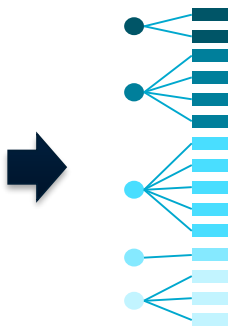
1. User defined Genomic constraints

5 x 1-way interaction
5 x 2-way interactions
5 x 3-way interactions
5 x 4-way interactions
5 x 5-way interactions
... n-way



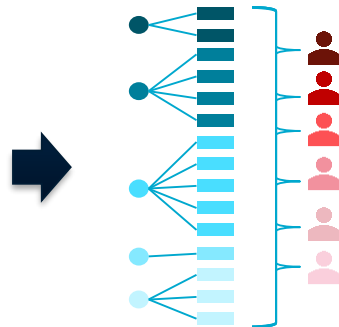
2. Random SNP selection

PEPS calculates the number of SNPs needed to conform to the user defined genomic constraints and randomly selects them from the VCF



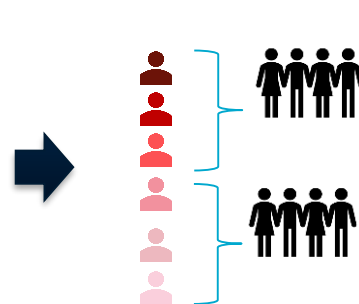
3. Variables are assigned a phenotypic effect

The sum of all partial effects result in the final phenotype



4. Individuals are scored and ranked

the cumulative effects of each variable is calculated for all samples



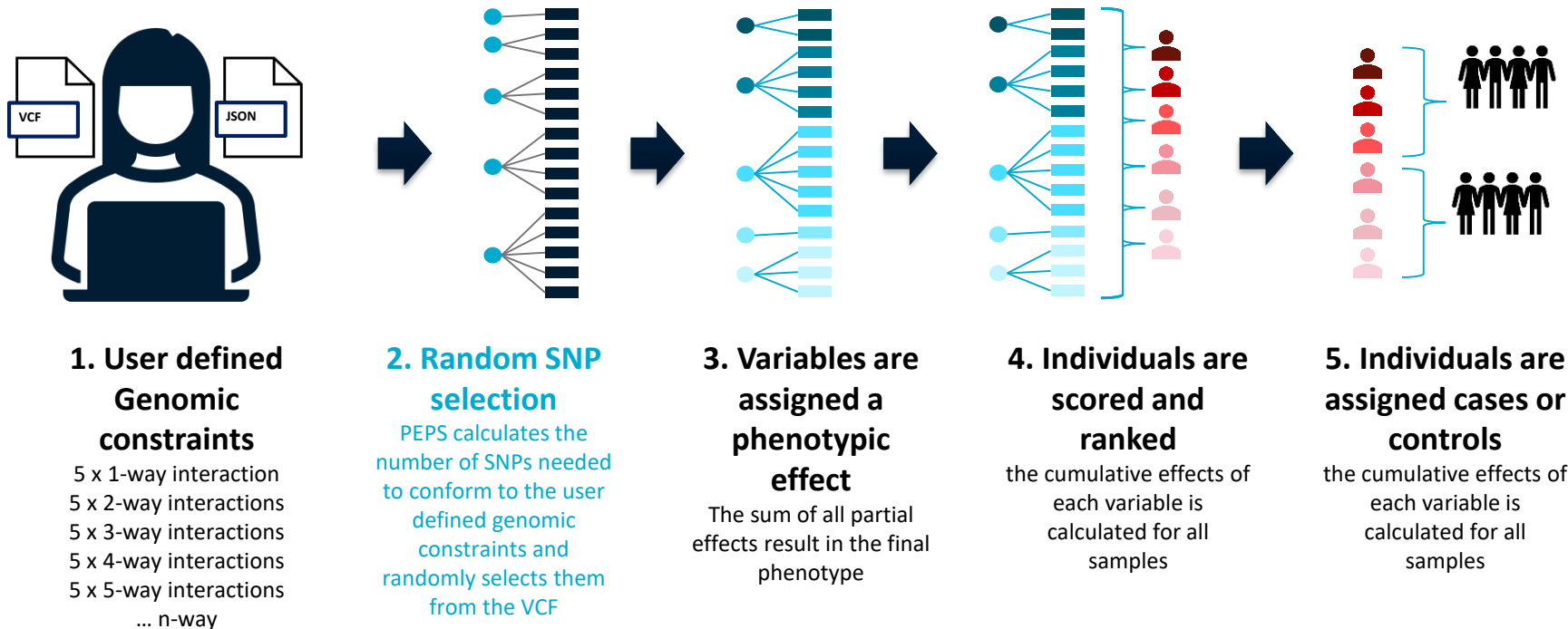
5. Individuals are assigned cases or controls

the cumulative effects of each variable is calculated for all samples

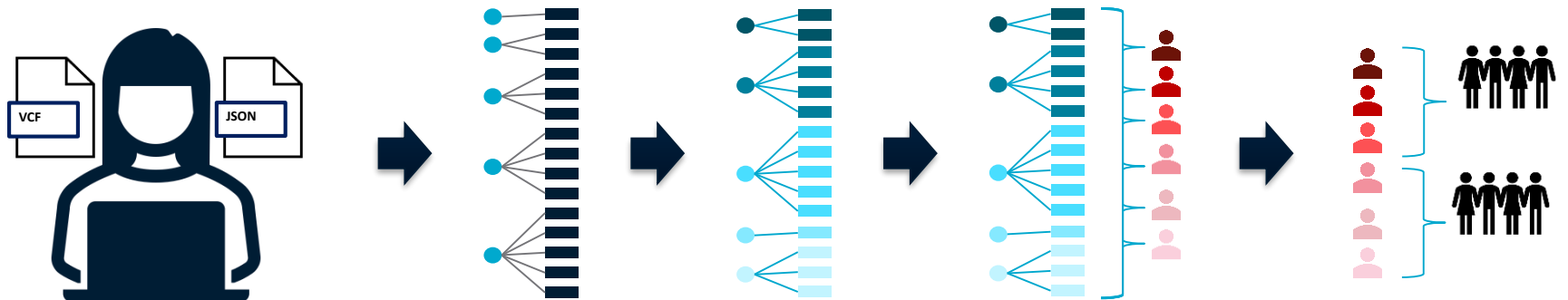


Australian e-Health
Research Centre

PEPS algorithm



PEPS algorithm



1. User defined Genomic constraints

5 x 1-way interaction
5 x 2-way interactions
5 x 3-way interactions
5 x 4-way interactions
5 x 5-way interactions
... n-way

2. Random SNP selection

PEPS calculates the number of SNPs needed to conform to the user defined genomic constraints and randomly selects them from the VCF

3. Variables are assigned a phenotypic effect

The sum of all partial effects result in the final phenotype

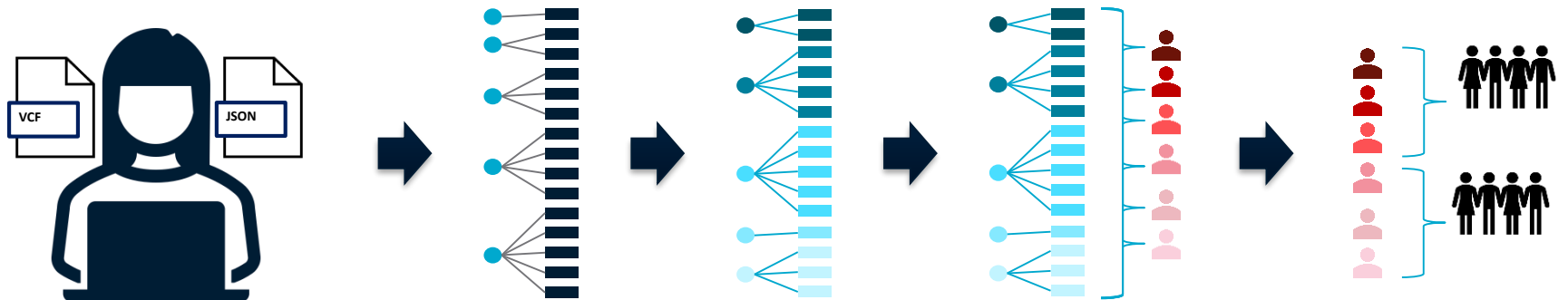
4. Individuals are scored and ranked

the cumulative effects of each variable is calculated for all samples

5. Individuals are assigned cases or controls

the cumulative effects of each variable is calculated for all samples

PEPS algorithm



1. User defined Genomic constraints

5 x 1-way interaction
5 x 2-way interactions
5 x 3-way interactions
5 x 4-way interactions
5 x 5-way interactions
... n-way

2. Random SNP selection

PEPS calculates the number of SNPs needed to conform to the user defined genomic constraints and randomly selects them from the VCF

3. Variables are assigned a phenotypic effect

The sum of all partial effects result in the final phenotype

4. Individuals are scored and ranked

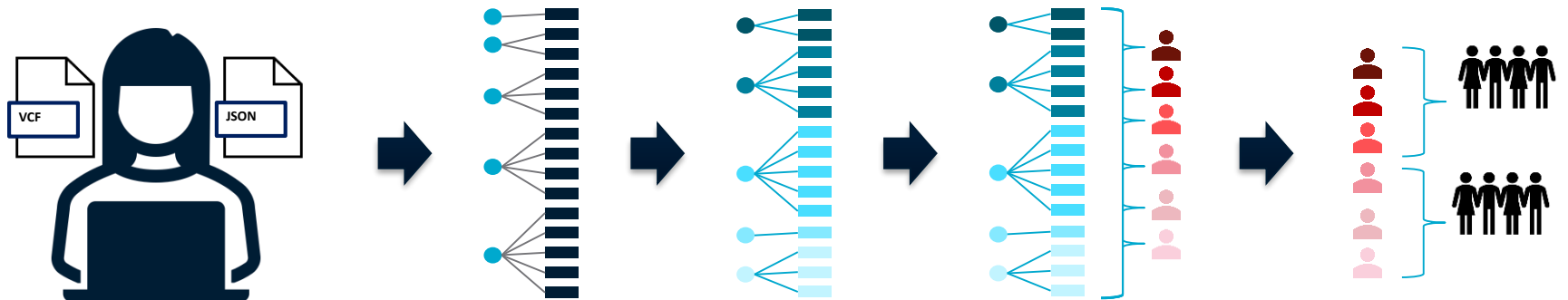
the cumulative effects of each variable is calculated for all samples

5. Individuals are assigned cases or controls

the cumulative effects of each variable is calculated for all samples

PEPS algorithm

PEPS outputs summary data on variants used to simulate the phenotype



1. User defined Genomic constraints

5 x 1-way interaction
5 x 2-way interactions
5 x 3-way interactions
5 x 4-way interactions
5 x 5-way interactions
... n-way

2. Random SNP selection

PEPS calculates the number of SNPs needed to conform to the user defined genomic constraints and randomly selects them from the VCF

3. Variables are assigned a phenotypic effect

The sum of all partial effects result in the final phenotype

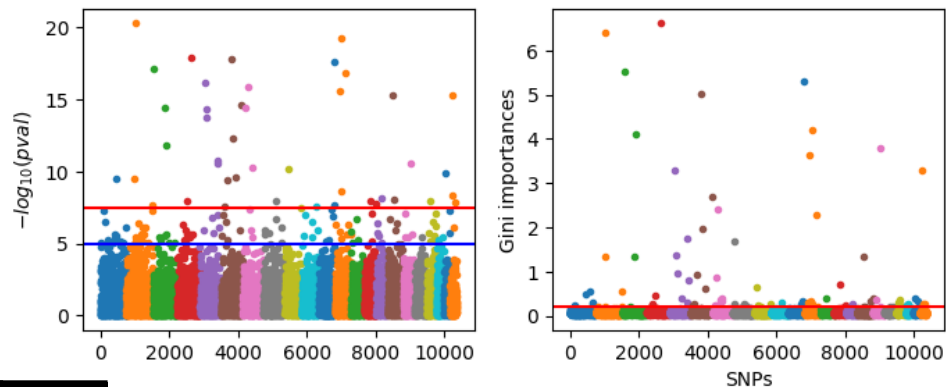
4. Individuals are scored and ranked

the cumulative effects of each variable is calculated for all samples

5. Individuals are assigned cases or controls

the cumulative effects of each variable is calculated for all samples

Example- 1000 genomes project



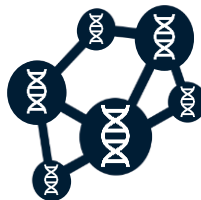
Epistatic order	Variables (n)	Truth SNPs (n)	Truth SNPs RF (n)	Truth SNPs RF (%)	Truth SNPs LR (n)	Truth SNPs LR (%)
1	5	5	3	60%	0	0%
2	5	10	8	80%	7	70%
3	5	15	12	80%	10	66%
4	5	20	11	55%	10	50%
5	5	25	15	60%	10	40%

SNPs used to simulate a phenotype were identified using a genome wide association analysis

PEPS



Share securely
without
compromising privacy



generate phenotypes
that emulates
epistatic interactions



Develop software
using known targets



Bypass data
collecting phase of
development –
accessing
consortium data

PEPS is a tool to support the development of genomic tools and pipelines



Australian e-Health
Research Centre



MED INFO23

8 – 12 JULY 2023 | SYDNEY, AUSTRALIA

Visit our site:

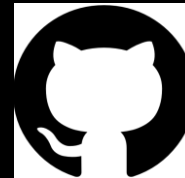
<https://bioinformatics.csiro.au>

@Tbioinf

Thank you

CSIRO Health and
Biosecurity
Mitchell O'Brien

Mitchell.O'Brien@csiro.au



PEPS is
on
Github



Australian e-Health
Research Centre