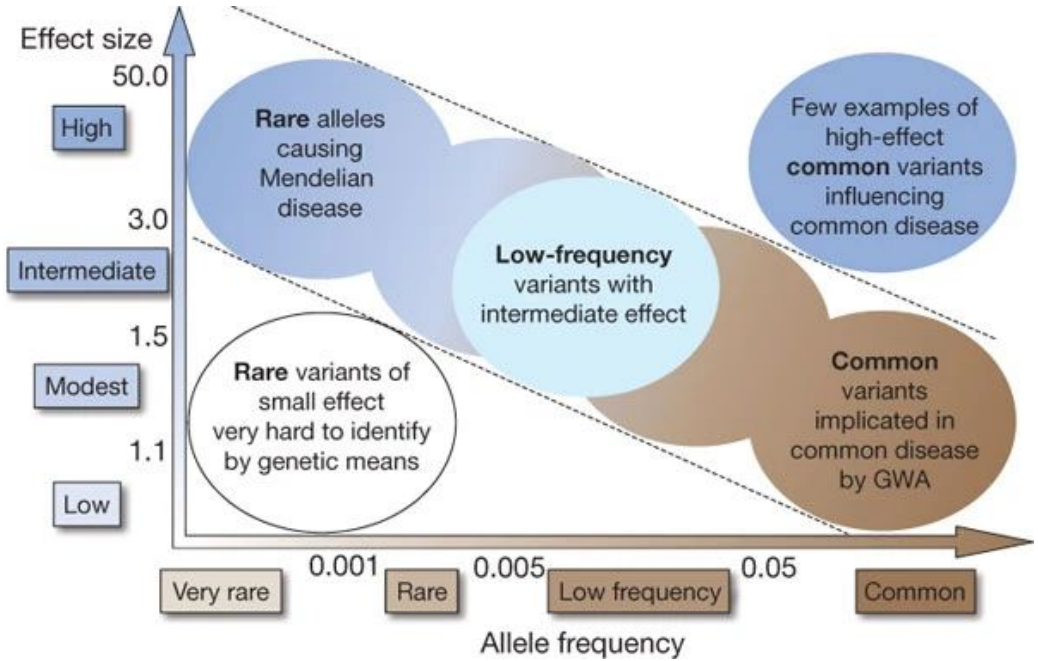
Cis and trans effects on variant penetrance

Scientific objective

Rare variants with large effect sizes (high penetrance)

Common (ancestral) variants typically have low effects and may affect disease risk through a different mechanism

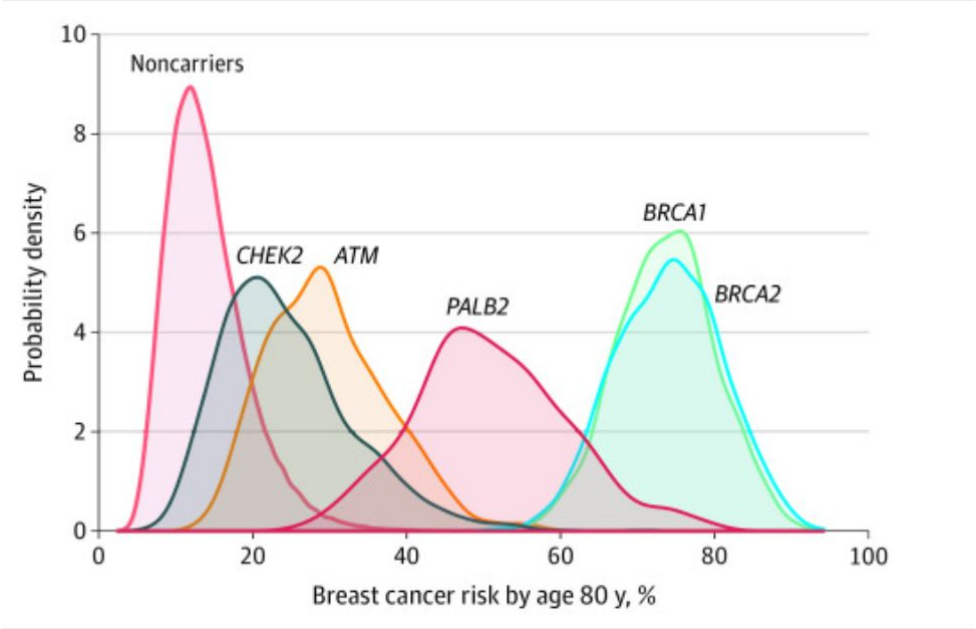
Objective: Detecting the interactions between highly penetrant effects and ancestral variants



Finding the missing heritability of complex diseases

Why look for interactions

- Personalized predictionions
  - moderate-risk genes (e.g. CHEK2, ATM, PALB2) and a low PRS brings risk down to non-non-carriers
- Disease mechanism discovery



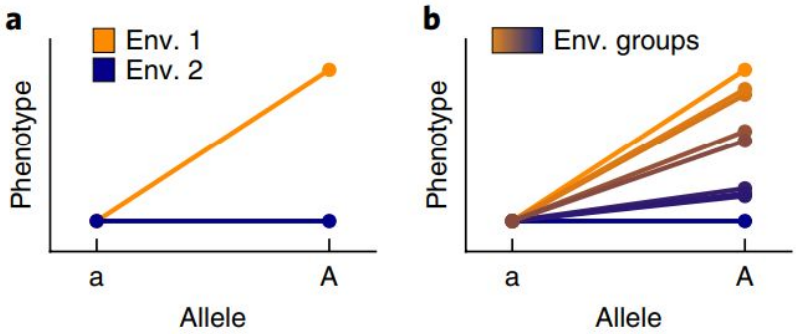
Modification of Lifetime Breast Cancer Risk for Pathogenic Variant Carriers and Noncarriers by an 86–Single-Nucleotide Variant Score.

Struct-LMM background

TECHNICAL REPORT
https://doi.org/10.1038/s41588-018-0271-0
nature genetics

A linear mixed-model approach to study multivariate gene–environment interactions

Rachel Moore[1,2,3,9], Francesco Paolo Casale[4,9], Marc Jan Bonder[2], Danilo Horta[2], BIOS Consortium[5], Lude Franke[6], Inês Barroso[1*] and Oliver Stegle[2,7,8*]



The environment is handled through the covariance and can be multivariate

Our “environment” is local genetic ancestry

The Struct-LMM model scales linearly with number of individuals

Struct-LMM Maths & Input Formats

$$y = M\alpha + g \odot \beta + e + \varepsilon,$$

	Phenotype
Individual 1	1
Individual 2	0
Individual 3	0
Individual 4	1

y

(the outcome)

	Dummy	PC1	PC2
Individual 1	1	0.1	0.9
Individual 2	1	0.9	-0.7
Individual 3	1	-0.3	0.1
Individual 4	1	0.1	0.2

M

(the covariates)

	SNV
Individual 1	1
Individual 2	0
Individual 3	2
Individual 4	0

g

(genetic variant)

	PC1	PC2
Individual 1	0.1	0.9
Individual 2	0.9	-0.7
Individual 3	-0.3	0.1
Individual 4	0.1	0.2

E

(haplotype configuration)

Usage and Results

```
python gxg-structlmm-script.py \
 --pcs output/PCoutput_local_ancestry_pcs.csv \
 --snv output/SNVoutput.txt  \
 --phenotype testPlink/synthetic_small_v1.pheno1 \
 --phenotype-column "Phenotype(binary)" \
 --output results.csv
```

```
           * * *

Tit   = total number of iterations
Tnf   = total number of function evaluations
Tnint = total number of segments explored during Cauchy searches
Skip  = number of BFGS updates skipped
Nact  = number of active bounds at final generalized Cauchy point
Projg = norm of the final projected gradient
F     = final function value

           * * *

   N    Tit     Tnf  Tnint  Skip  Nact     Projg        F
    2      8     13      8     0     0   3.247D-04   8.503D+02
  F =   850.26302680069568

CONVERGENCE: REL_REDUCTION_OF_F_<=_FACTR*EPSMCH

Results:
P-value: 8.64303433e-01
Saving results to results.csv
```

Next steps

- Run on actual data with real phenotypes
- Write wrappers to test all recombination blocks
- Find some significant interactions