ROSLIN

ASSOCIATION ANALYSES of the MAS-QTL DATA SET using GRAMMAR, PRINCIPAL COMPONENTS and BAYESIAN NETWORK METHODOLOGIES

Burak Karacaören, Tomi Silander, José M. Álvarez-Castro, Chris S. Haley, Dirk Jan de Koning

ROSLIN INSTITUTE and R(D)SVS, UNIVERSITY of EDINBURGH
SCOTLAND, UK

EBRC Easter Bush Research Consortium
BBSRC bioscience for the future

Polonezköy

From Wikipedia, the free encyclopedia

Coordinates 41°06'36"N

This article **does not cite any references or sources**.
Please help improve this article by adding citations to reliable sources. Unsourced material may be challenged and removed. *(February 2007)*

"Adampol" redirects here. For other uses, see Adampol (disambiguation).

Polonezköy or **Adampol** is a small village at the Asian side of Istanbul, about 30 kilometers away from the historic city center, within the boundaries of the Beykoz district. It was founded in 1842 by Polish settlers.


Our Lady of Częstochowa Chu

Contents [hide]

1 History of Polonezköy
2 Monuments and places of interest
3 See also
4 External links

History of Polonezköy [edit]

Polish-Turkish relations have been good since the 18th century, and the Ottoman Empire was the only major power in the world which never recognized the dissolution and partitioning of Poland between Austria-Hungary, Russia and Prussia, while Constantinople (Istanbul) remained as the only capital city in the world to maintain a "Polish Ambassador" until the end of the First World War and the subsequent re-creation of Poland.

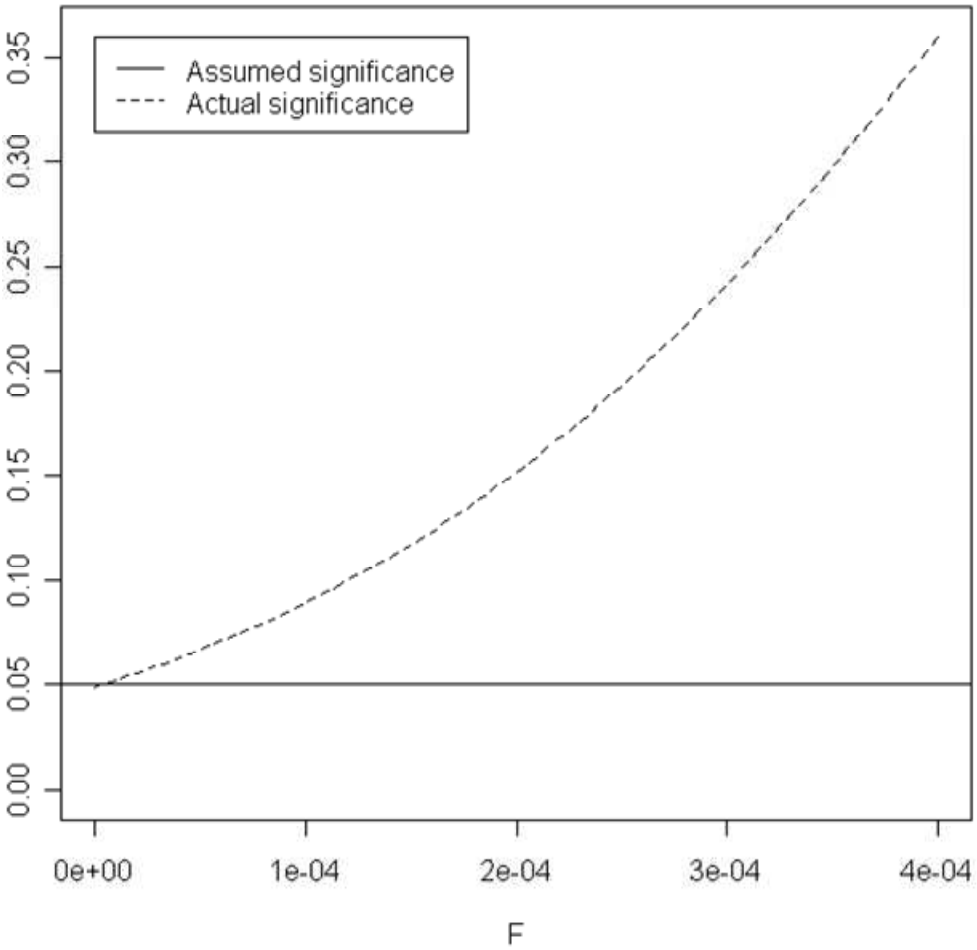
Polonezköy (Adampol) was founded by Duke Adam Czartoryski in 1842. He was the Chairman of the Polish National Uprising Government and the leader of a political emigration party. The settlement was named Adam-koj (Adamköy) after its founder, which means the "Village of Adam" in Turkish (*Adampol* means "Town of Adam" in Polish).

Duke Adam Czartoryski wanted to create the second emigration centre here (the first one was in Paris, France.) He sent his representative, Michał Czajkowski, to Turkey. Michał Czajkowski, after converting to Islam in 1850, became known as Mehmed Sadyk Pasza (Mehmet Sadık Paşa). He purchased the forest area which encompasses present-day Adampol from a missionary order of Lazarists. Plans were made to establish Adampol on this area in the future.

At the beginning, the village was inhabited by 12 people, but there were no more than 220 people when the village was most populated. In the

k1

In GWAS, genetic relations between and within cases and controls need to be taken into account



500 cases & 500 controls

10.000 markers

- **For isolated population with few founders problem increases.**

- **Possible Solutions**
  - **GRAMMAR(Aulchenko et al, 2007)**
  - **STRAT (Price et al, 2006)**

Fig1. The actual significance level of the allele test as a function of the coefficient of inbreeding.

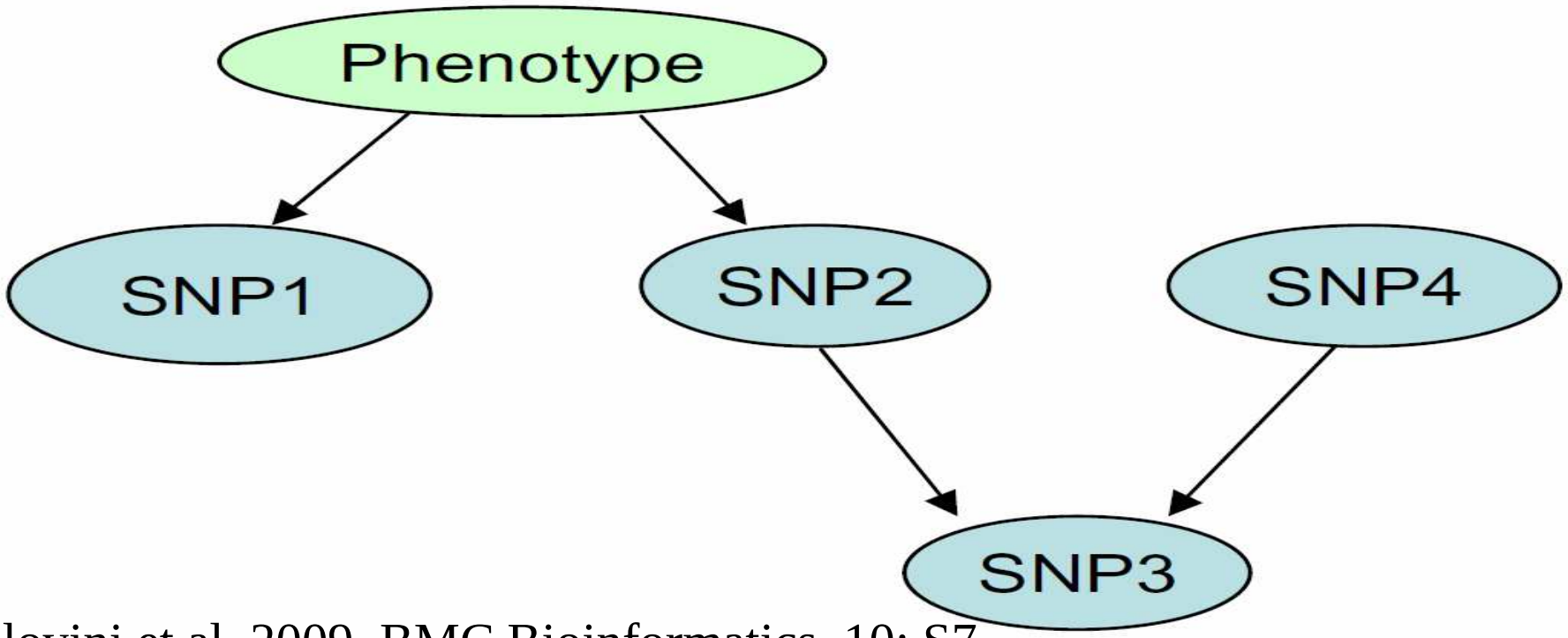
Siegmund and Jakir, 2007, The Statistics of Gene Mapping, Springer

Slajd 3

k1 Her ikisi de onemlilik, ama akrabaligin artmasina dayali olarak seviye dusuyir, boylece diyelim bir isaretci 0.10 cikiyor onemlilik, akrabalik arttikca buna onemlidir demen gerekiyor.
karacao; 2010-05-13

LINKAGE DISEQUILIBRIUM IN TOP SNPs FROM GWAS

- Many significant SNPs from GWAS in LD
  - Co-linearity among these SNPs.
  - Wang and Abbott (2008) suggested using (PCReg) approach.



Malovini et al, 2009, BMC Bioinformatics, 10: S7

8.47 x 11.00 in

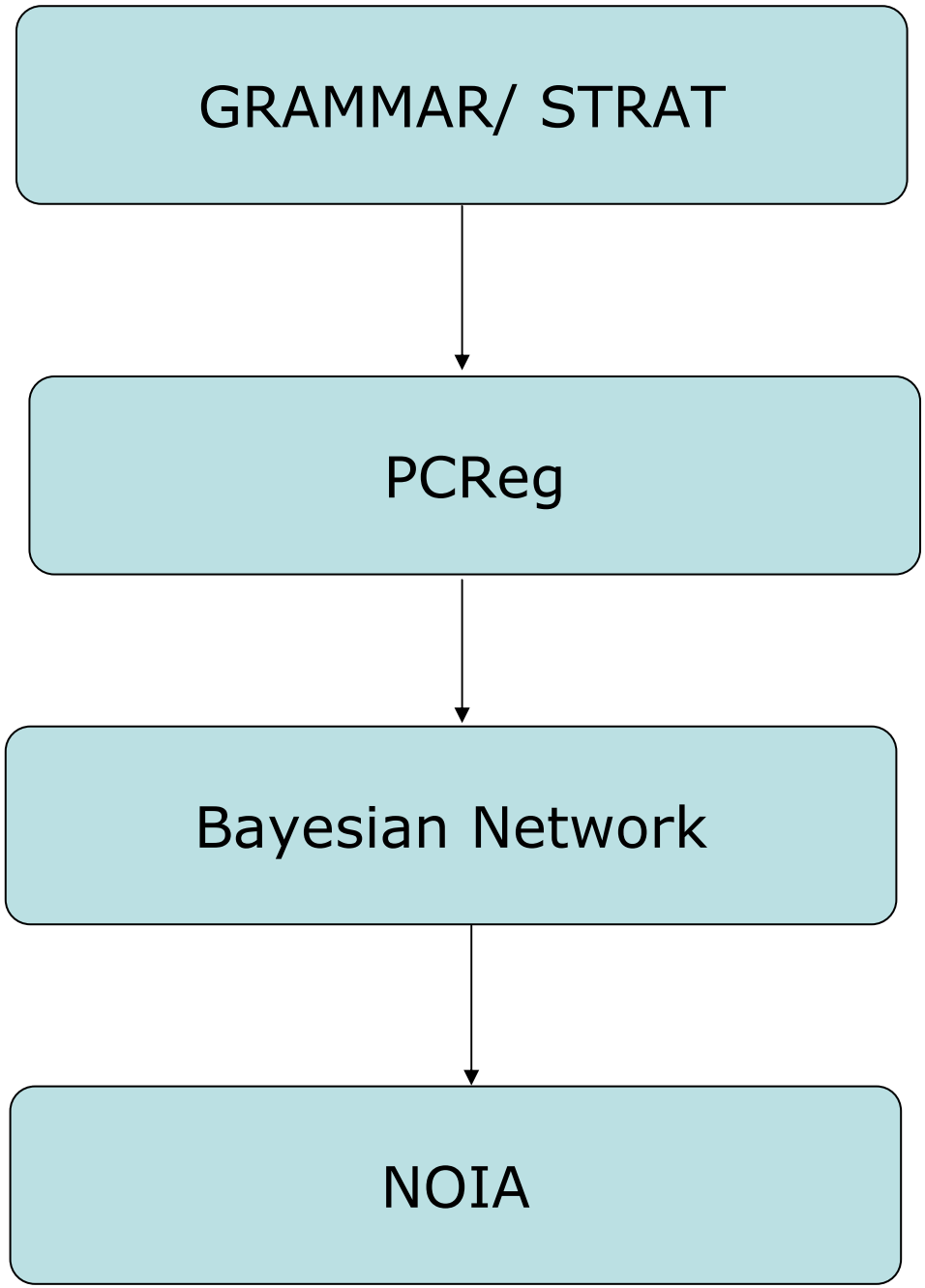
- **Most GWAS studies do not model correlations:**
  - **Among SNPs**
  - **Between SNPs and environmental variables**

Bayesian networks are models that present statistical dependencies and independencies in the joint probability distribution of the data.

ESTIMATION of SNPs Effects

Residuals may be used, but these may not be normally distributed.

The Natural and Orthogonal interactions model (NOIA) (Castro et al, 2008) could be useful to decompose effects orthogonally.


GRAMMAR/ STRAT
PCReg
Bayesian Network
NOIA

Genome-wide Rapid Association using Mixed Model and Regression

$$y = \mathbf{X}b + \mathbf{Z}a + e \quad (1)$$

$$\mathbf{y} = \mathbf{Xb} + \boldsymbol{\eta} + \mathbf{e} \quad (2)$$

Principal component analyses. Principal components analyses used to orthoganize the genomic space;

$$\begin{aligned} y_1 &= a_{11}x_1 + a_{12}x_2 + \ldots + a_{1p}x_p \\ y_2 &= a_{21}x_1 + a_{22}x_2 + \ldots + a_{2p}x_p \\ &\vdots \\ y_p &= a_{p1}x_1 + a_{p2}x_2 + \ldots + a_{pp}x_p \end{aligned}$$

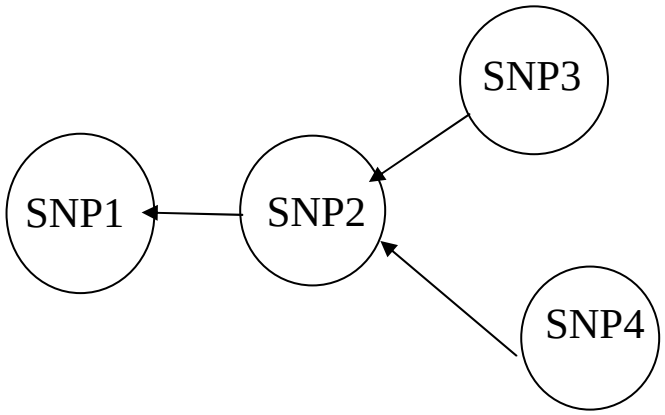
with the coefficients being chosen so that $y_1, y_2, \ldots, y_p$ account for most of the explanatory proportions of the total variance of the original variables, $x_1, x_2, \ldots, x_p$, (Everitt et al., 2001).

BAYESIAN NETWORK

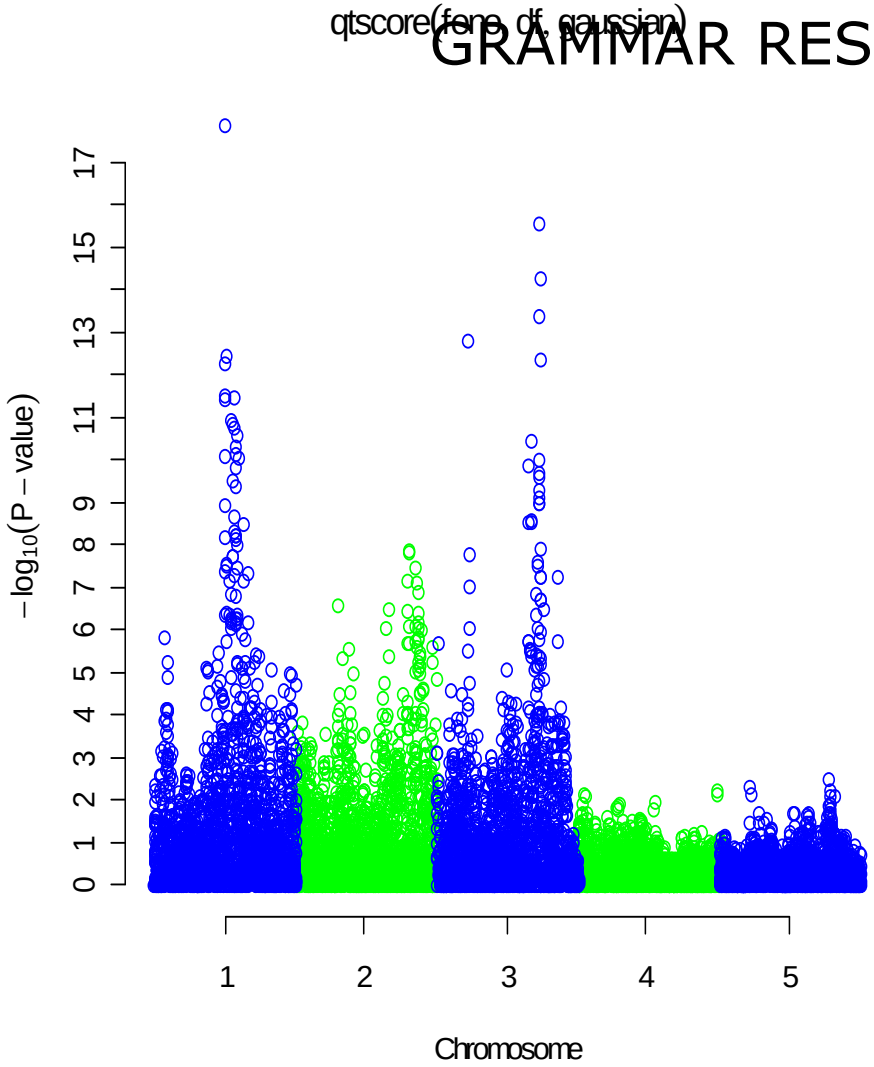
An important property of Bayesian network models is that the joint probability distribution over the model variables factorizes to a product of *n* conditional probability distributions:

$$P(X_1,\ldots,X_n)=\prod_{i=1}^{n}P\left(X_i\middle|\Pi_i\right),$$

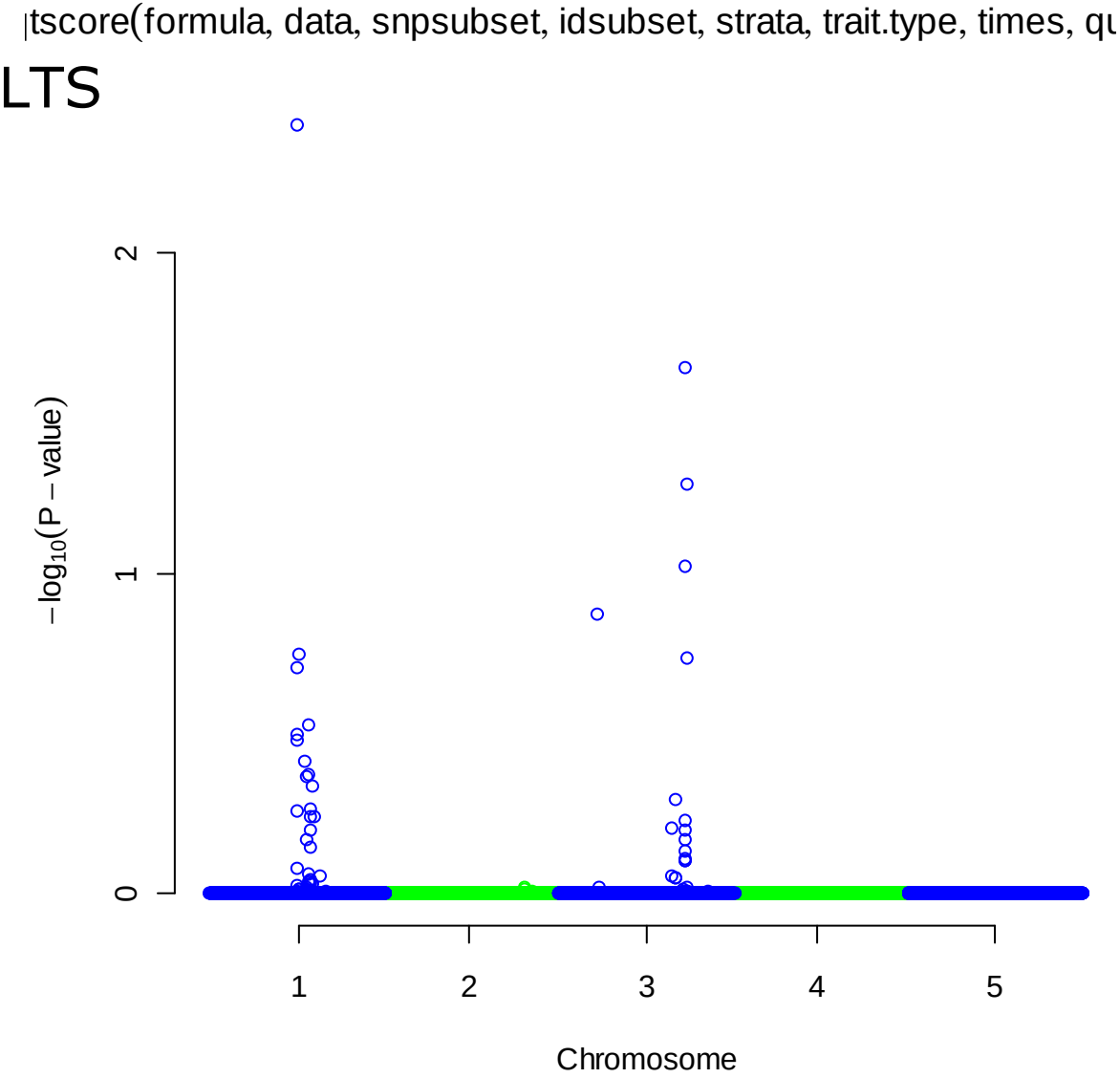
where Π_i denotes the parents of variable X_i (Myllymaki etal, 2002).



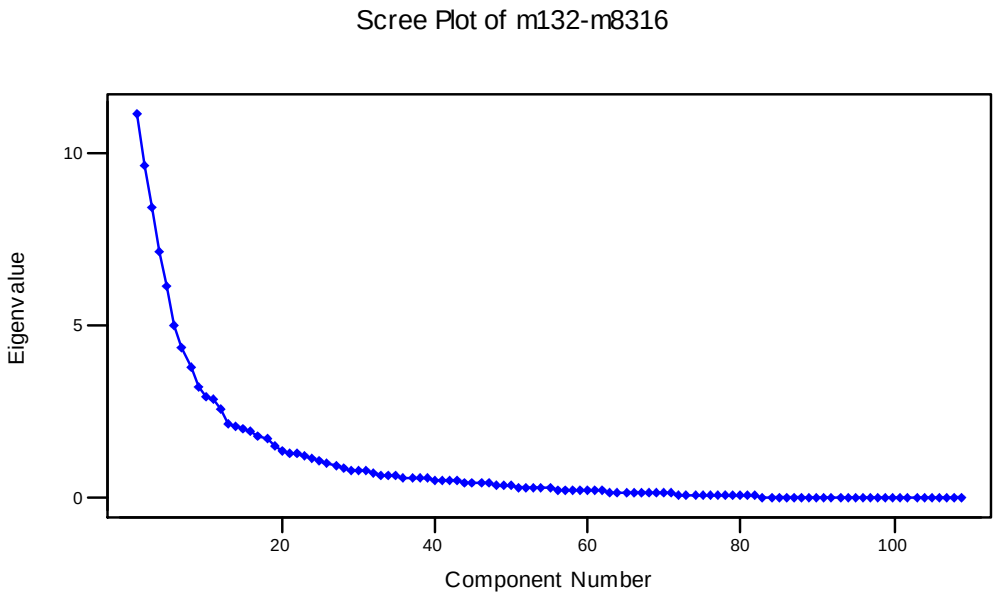
GRAMMAR RESULTS



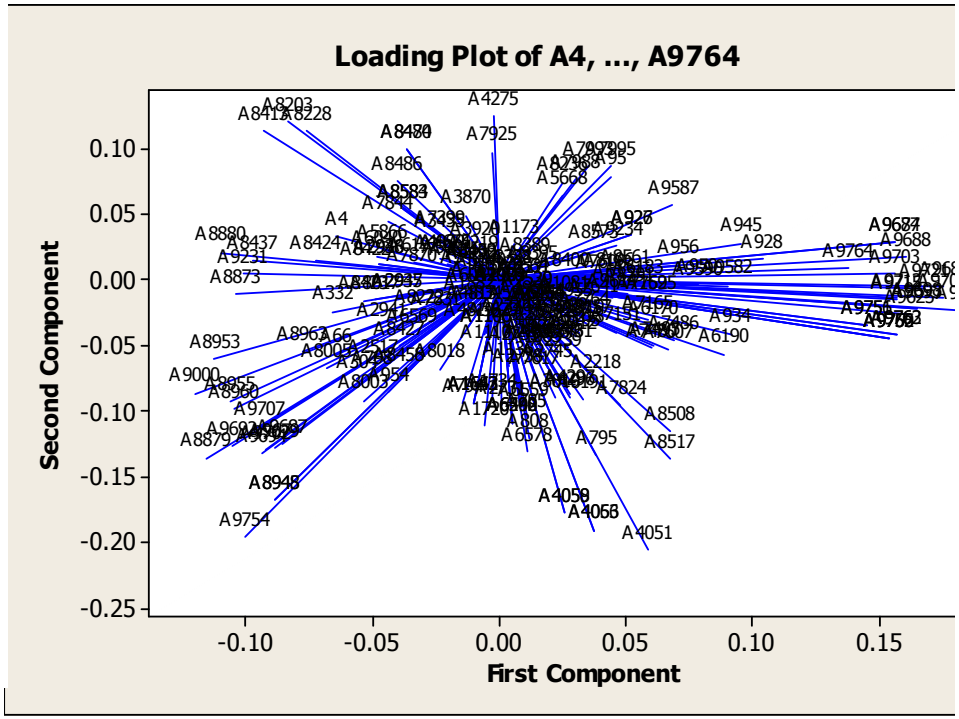
Before permutation



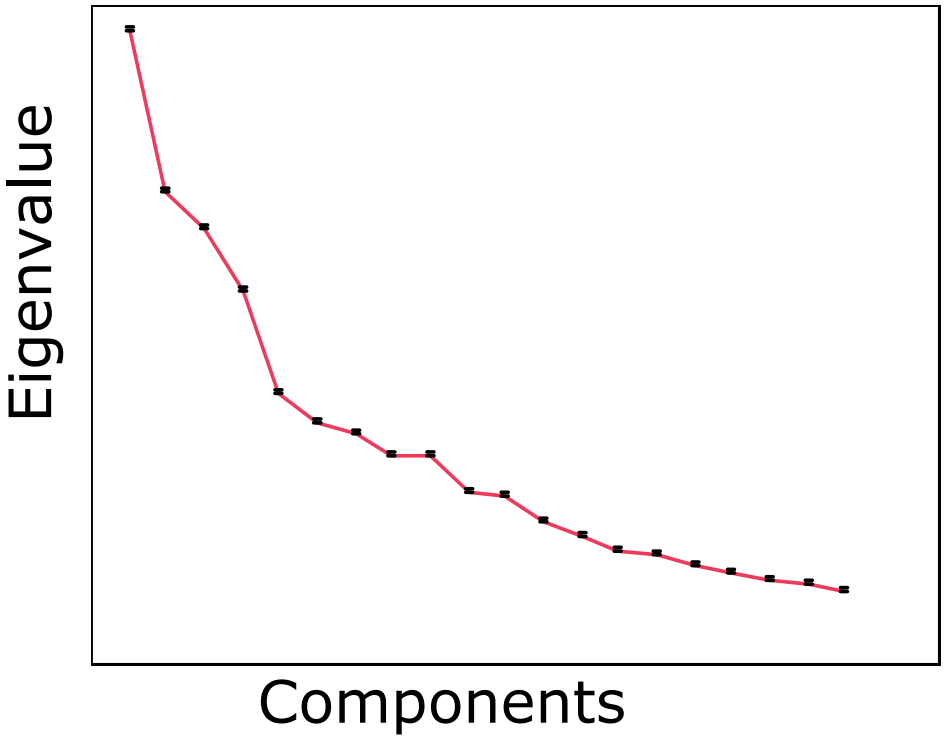
After 1000 permutations



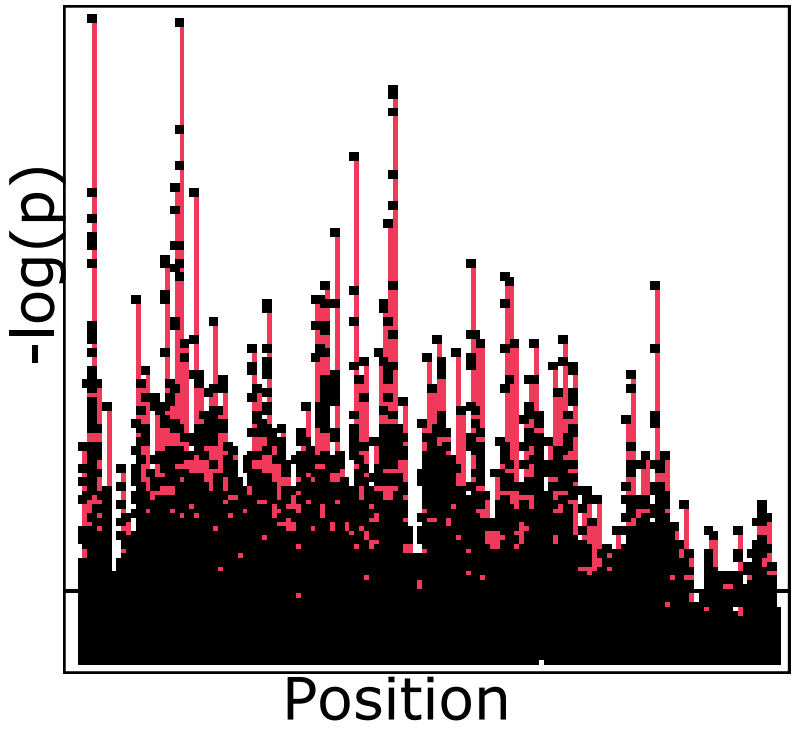
Scree plot from principal components of Top SNPs based on GRAMMAR.



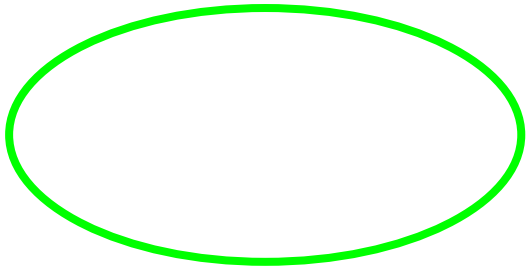
Loading plot for first 2 principal components.

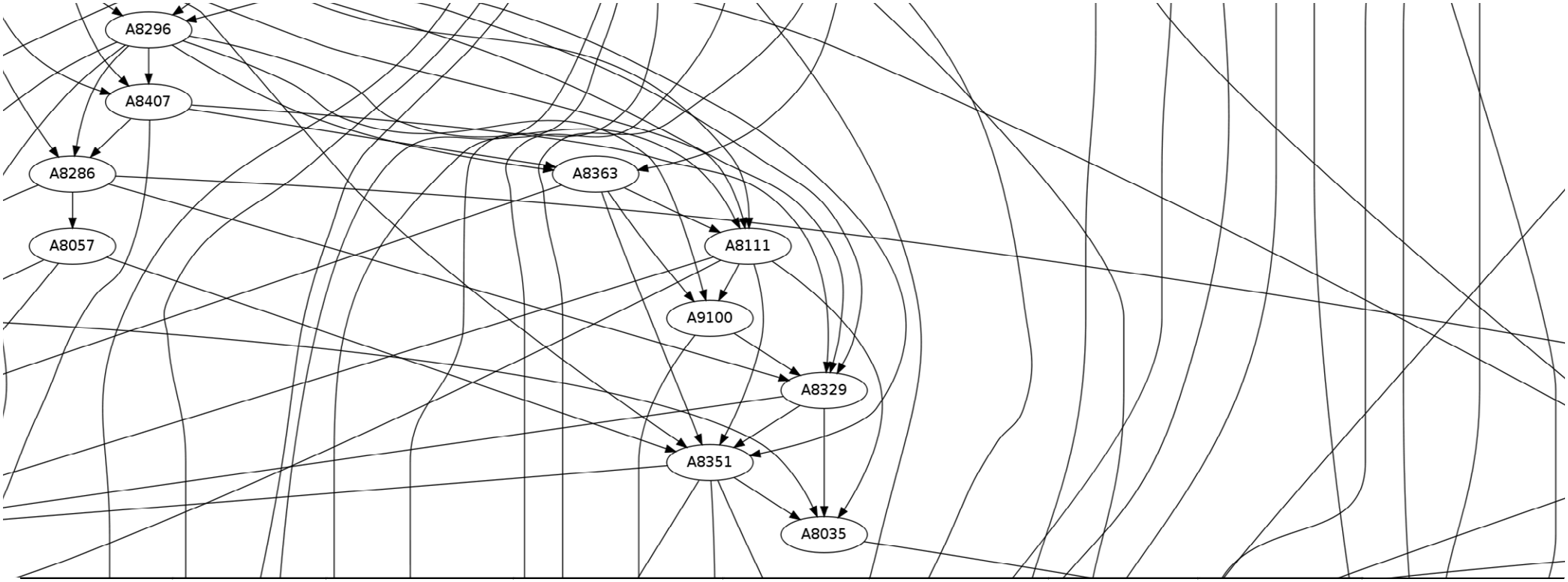


Scree plot using 20 principal components for Binary trait.

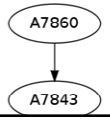


Results of principal component stratification based on 10 principal components.

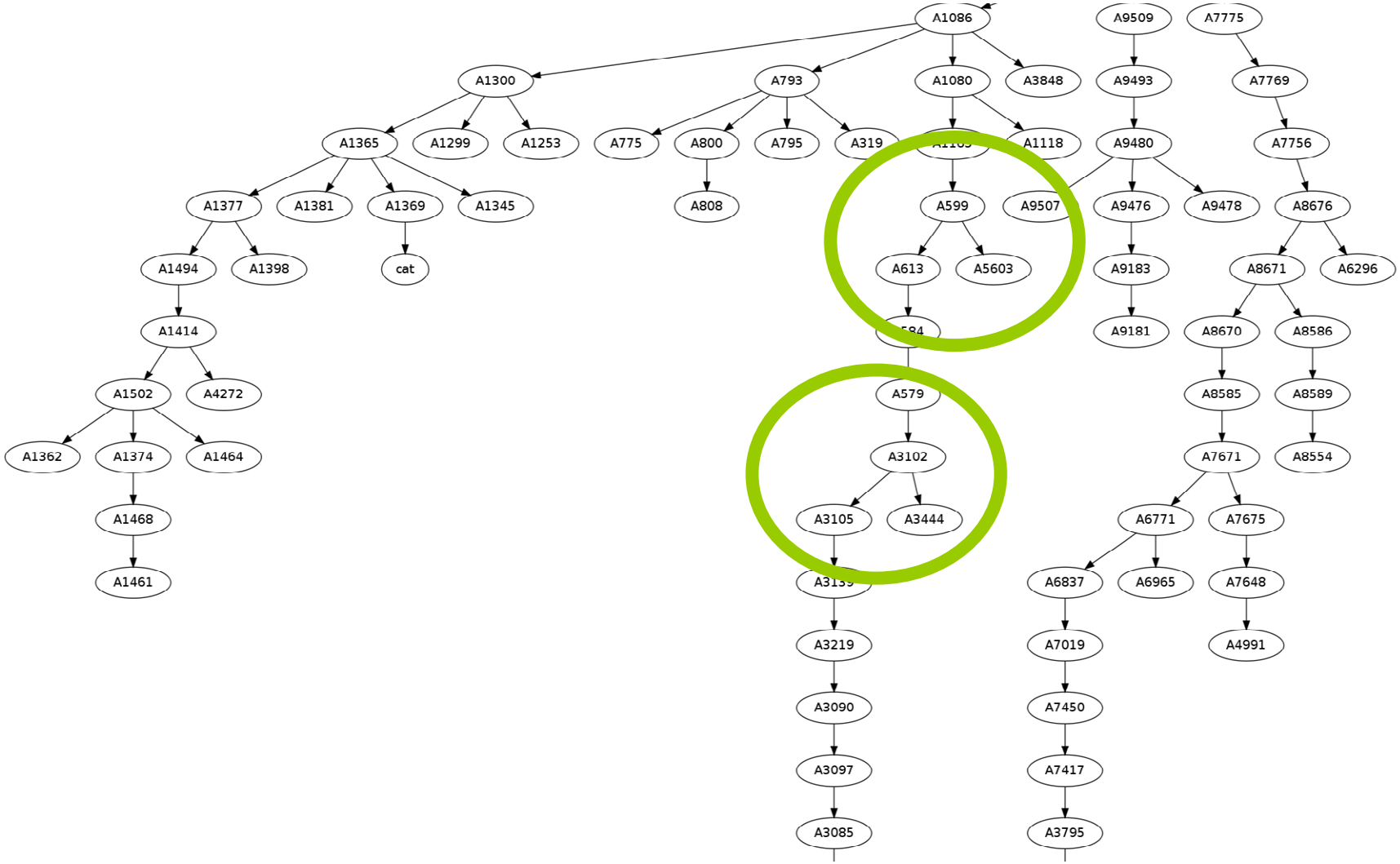




Marker1	Marker2	Chi	P(Chi)	D	CorrCoeff	Dprime	ARC	exp(arc)
A8111	A9100	692.1253	1.50E-152	0.08924	0.54643	0.86642	293.5266	2.999E+127
A8363	A9100	516.7682	2.10E-114	0.08302	0.47216	0.98559	71.6831	1.3539E+31
A8111	A8363	548.7223	2.40E-121	0.11081	0.48654	0.64052	519.4383	3.883E+225
A8111	A8351	2318	1.70E-236	0.15661	0.68207	0.97544	694.4988	4.14E+301
A8035	A8329	1668.985	0	0.20572	0.84853	0.97144	232.6614	1.105E+101
A8329	A8351	20.12108	7.27E-06	-0.0232	-0.0932	-0.1084	240.8512	3.984E+104



Marker1	Marker2	ChiSq	ProbChi	D	CorrCoeff	Dprime	Delta	PropDiff	YulesQ	ARC
A599	A613	1567.399	0	0.09457	0.82231	0.95048	0.74373	0.73941	0.99602	850.271
A599	A5603	117.2745	2.50E-27	0.03616	0.22493	0.6041	0.15173	0.1447	0.65856	80.5
A3102	A3105	1916.852	0	0.19996	0.90936	1	1	0.9396	1	1668.336
A3102	A3444	128.9518	6.95E-30	0.02766	0.23586	0.65655	0.67148	0.45698	0.76186	80.14



	Linear Model		NOIA	
	Residual(%)	Phenotype(%)	Residual(%)	Phenotype(%)
First100SNPs	14.2	35.2	16.02	200.46
Random100SNPs	2.9	11.8	1.45	50.42

Table 1. Estimation of SNPs effects for first 100 and random 100 SNPs with linear and NOIA model.

SNP	LINEAR(%)	NOIA(%)
1	0.8	0.8
2	0.1	0.08
3	1.0	0.99
4	1.3	1.34
5	0.3	0.3
6	0.1	0.1
7	0.3	0.3
8	0.7	0.7
9	2.9	2.9
10	0.1	0.1
11	0.1	0.3
12	0.0	0.02
13	0.3	0.4
14	0.1	0.2
15	0.4	0.5
16	0.0	0.001

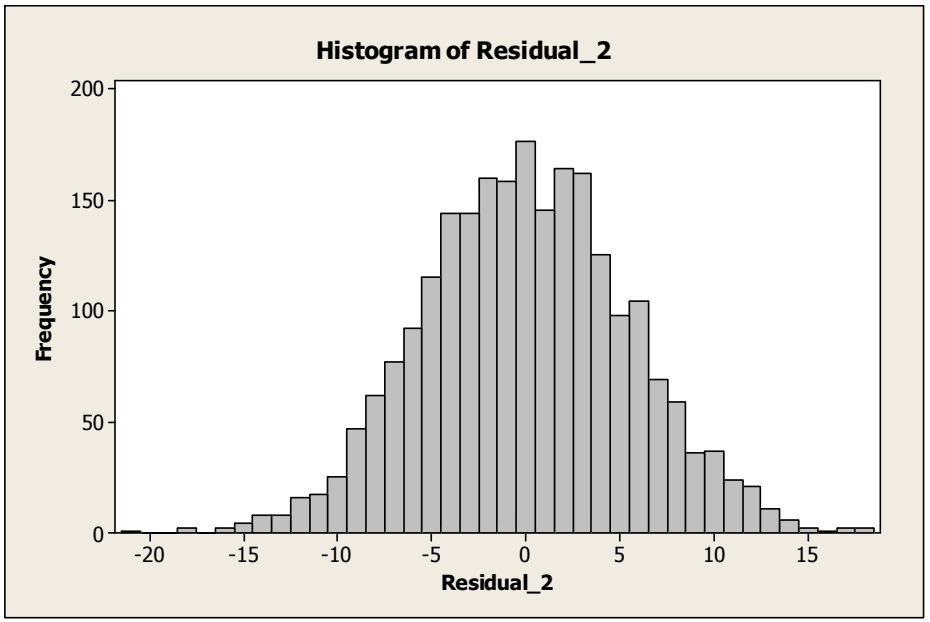


Table 2. Estimates of explanatory proportions for top SNPs from linear and NOIA models.

DISCUSSIONS/CONCLUSIONS

- GRAMMAR can be used to accommodate genetic relationship among cases and controls in GWAS.
- In practice NOIA could be used to predict SNPs effects with dominant effects.
- PCReg is useful to choose most important SNPs from list of SNPs in linkage disequilibrium.
- Bayesian Tree Structured Networks useful to introduce and investigate relationships within/ among SNPs and other environmental effects.

ROSLIN

REFERENCES

Myllymaki, P., Silander, T., Tirri, H., Uronen, P. B-Course: A Web-Based Tool for Bayesian and Causal Data Analysis. (2002) *International Journal on Artificial Intelligence Tools*, Vol 11, No. 3, 369-387.

Siegmund D. and Yakir, B. (2007) The statistics of gene mapping. Springer.

Wang, K., and Abbott, D. (2008) A principal component approach to multilocus genetic association studies. *Genetic Epidemiology* 32:108-118.

- THANKS FOR YOUR ATTENTION

The natural and orthogonal inter-actions (NOIA) model (Álvarez-Castro and Carlborg 2007) is a genotype-to-phenotype map that unifies the so-called functional and statistical approaches (see Hansen and Wagner 2001). Here, we are interested in the statistical formulation of NOIA for two alleles, a parameterization of the expected phenotype of each genotype, *i.e.* the genotypic values G=(G_{ij}), in terms of genetic effects (additive and interaction effects). For one locus, this parameterization is **G**=**S**·**E** expanding to:

$$\begin{pmatrix} G_{11} \\ G_{12} \\ G_{22} \end{pmatrix} = \begin{pmatrix} 1 & -2p_2 & -\frac{p_{12}p_{22}}{2p_1p_2 - \frac{1}{2}p_{12}} \\ 1 & p_1 - p_2 & \frac{p_{11}p_{22}}{p_1p_2 - \frac{1}{4}p_{12}} \\ 1 & 2p_1 & -\frac{p_{11}p_{12}}{2p_1p_2 - \frac{1}{2}p_{12}} \end{pmatrix} \cdot \begin{pmatrix} \mu \\ \alpha \\ \delta \end{pmatrix}$$

The vector of genetic effects, **E**, entails the population mean, μ, the additive, α, and the dominance, δ, effects. The parameterization is given by the genetic-effect design matrix, **S**, expressed as a function of the genotypic frequencies of the population