

Max Planck Institute  
of Psychiatry

German Research Institute of Psychiatry



# BMM Group Meeting

**Ilaria Bonavita**

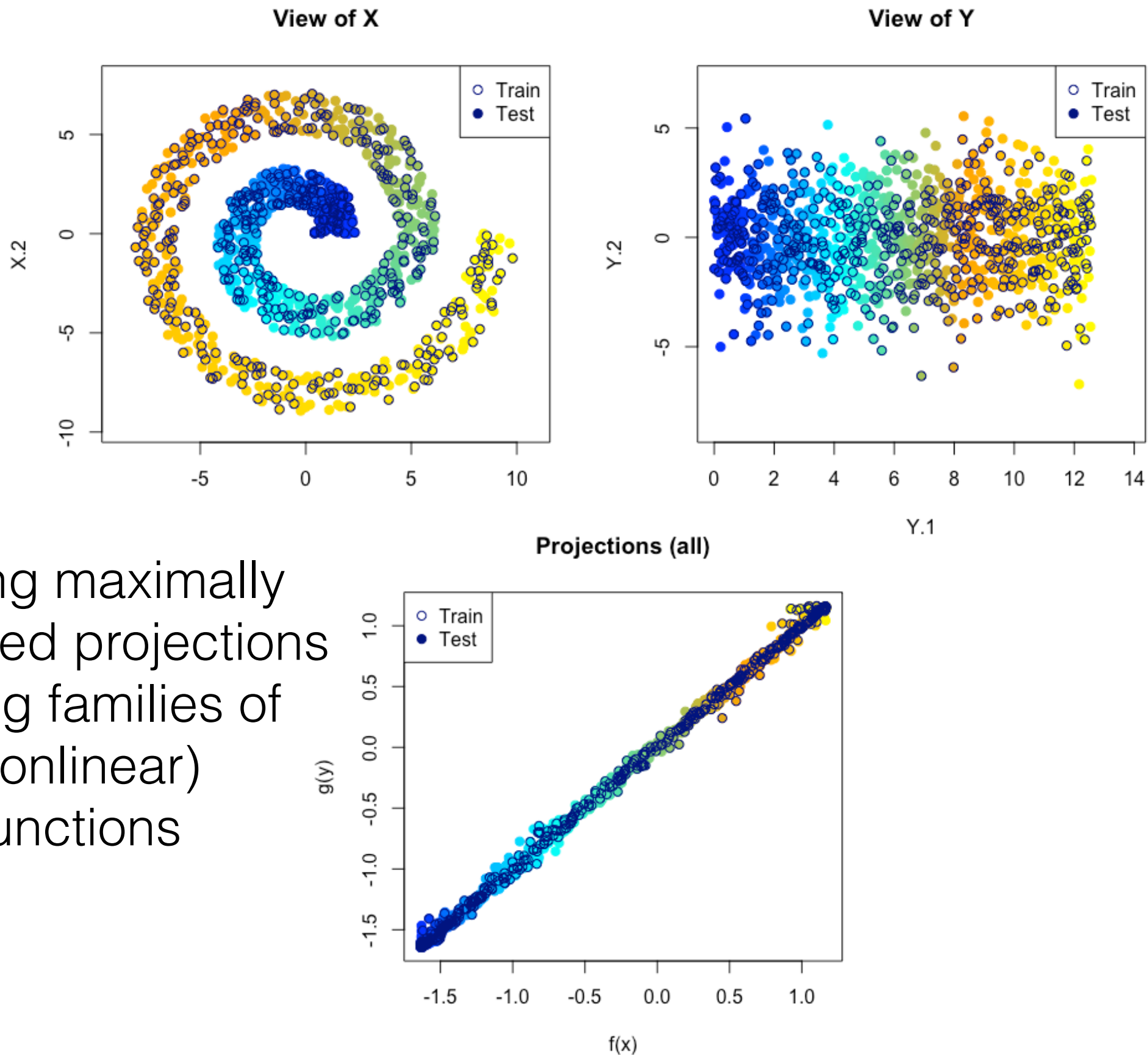
[ilaria\\_bonavita@psych.mpg.de](mailto:ilaria_bonavita@psych.mpg.de)

May 30, 2017

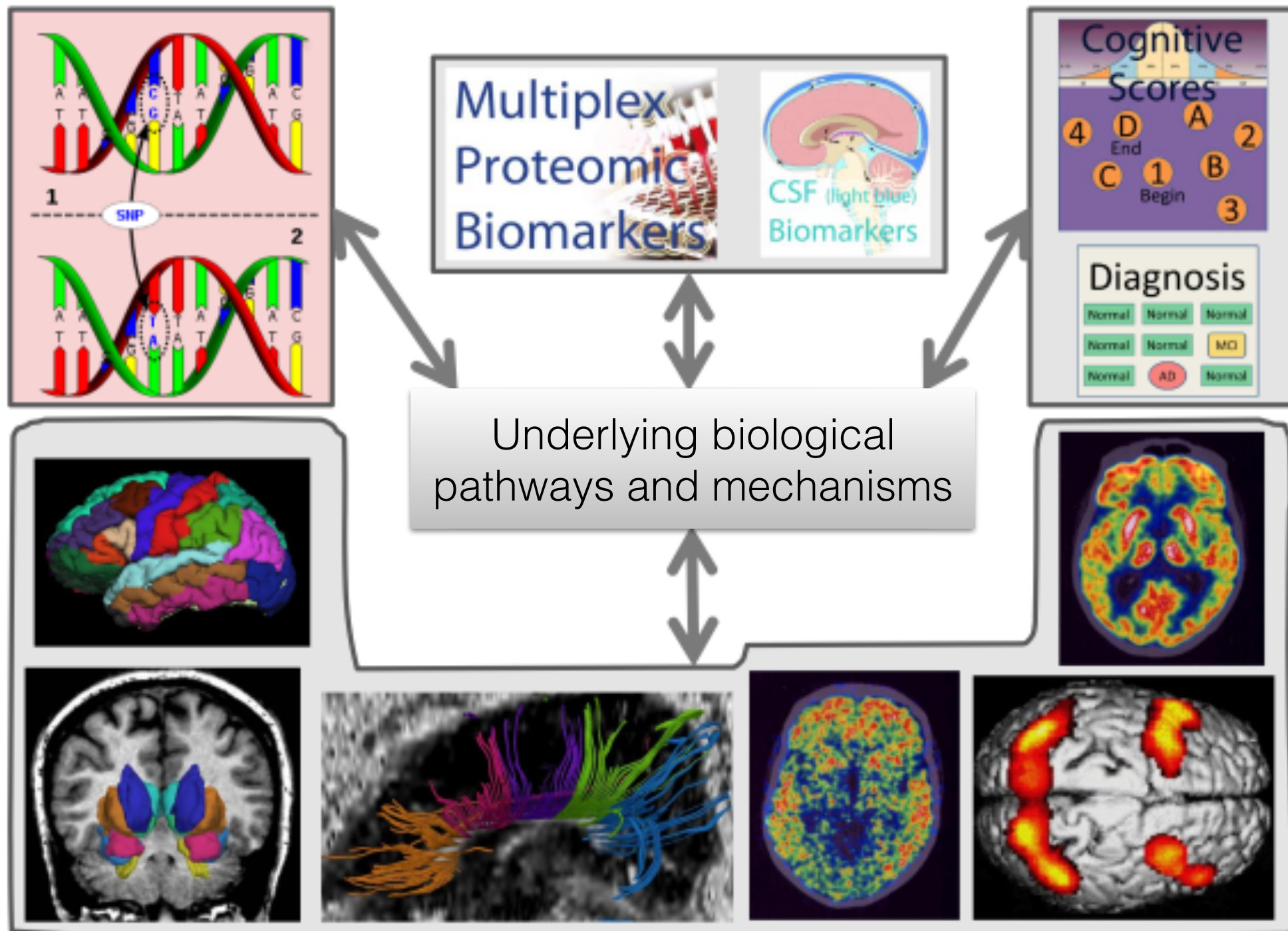
# Outline

- ▶ Background: review of Nonparametric canonical correlation analysis.
- ▶ Experiment: genotype - multi phenotype association in dyslexia data.
- ▶ Future steps.

# NCCA: a recap



Finding maximally  
correlated projections  
among families of  
(nonlinear)  
functions



high-dimensional ***X***

high-dimensional ***Y***



### **NCCA**

- dim of the transformed feature space ***d***
- kernel
- Number of nearest neighbours NNs
- ...



***d*** canonical correlation coefficients (+ p values)

d-dimensional (transformed)  
***X<sub>new</sub>***

d-dimensional (transformed)  
***Y<sub>new</sub>***

# Genotype - multi-phenotype correlation in dyslexia data

## **Predictors of developmental dyslexia in European orthographies with varying complexity.**

Landerl K<sup>1</sup>, Ramus F, Moll K, Lyytinen H, Leppänen PH, Lohvansuu K, O'Donovan M, Williams J, Bartling J, Bruder J, Kunze S, Neuhoff N, Tóth D, Honbolygó F, Csépe V, Bogliotti C, Iannuzzi S, Chaix Y, Démonet JF, Longeras E, Valdois S, Chabernaud C, Delteil-Pinton F, Billard C, George F, Ziegler JC, Comte-Gervais I, Soares-Boucaud I, Gérard CL, Blomert L, Vaessen A, Gerretsen P, Ekkebus M, Brandeis D, Maurer U, Schulz E, van der Mark S, Müller-Myhsok B, Schulte-Körne G.

- Developmental dyslexia affects 3-7% of the population.
- Mix of *biological* and *environmental* factors.
- *Orthography* has a central role in reading acquisition and dyslexia.
- Considerable variability in *symbol-sound consistency* among different orthographies.
- Explore the relation between orthography complexity and *cognitive mechanisms* underlying dyslexia.
- Large-scale cross-linguistic study (English, French, Dutch, German, Hungarian, Finnish).



# Data and preprocessing

Country	Cases	Controls	tot
Austria+Germany+Switzerland	1047	407	1454
Finland	161	163	324
Holland	136	148	284

SNPs
5747699
6197707
5978155

+ 10 Phenotype  
DS digit span; WR  
word reading; WS word  
spelling; NWR non-  
word reading; LRAN  
letters rapid automatic  
naming etc.

LD pruning - threshold: 0.2

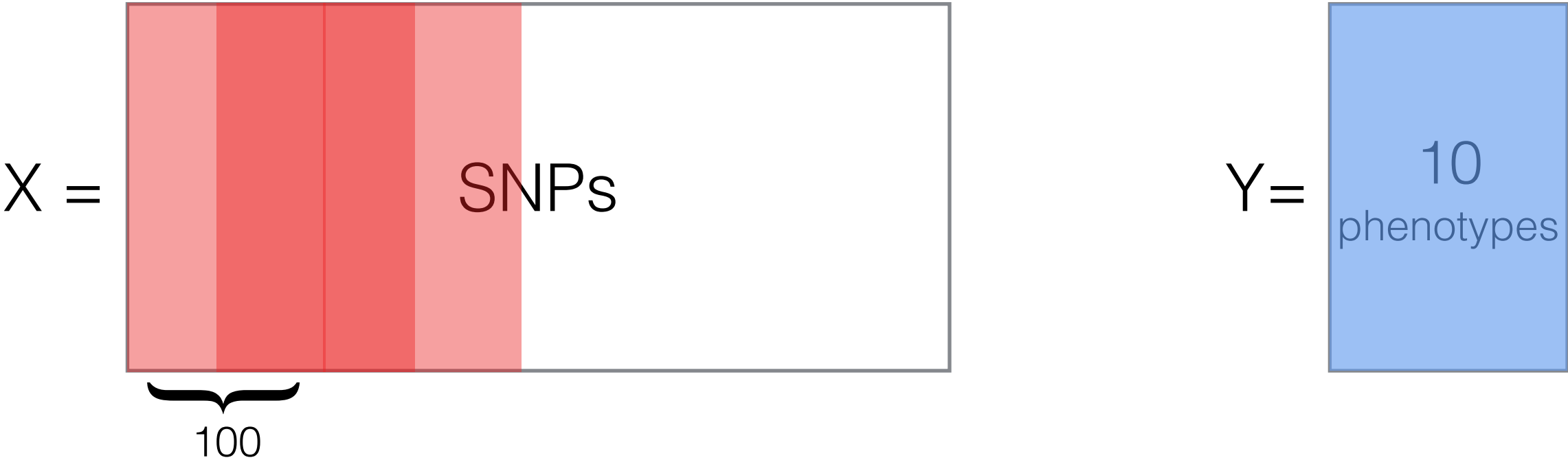


Dosage conversion

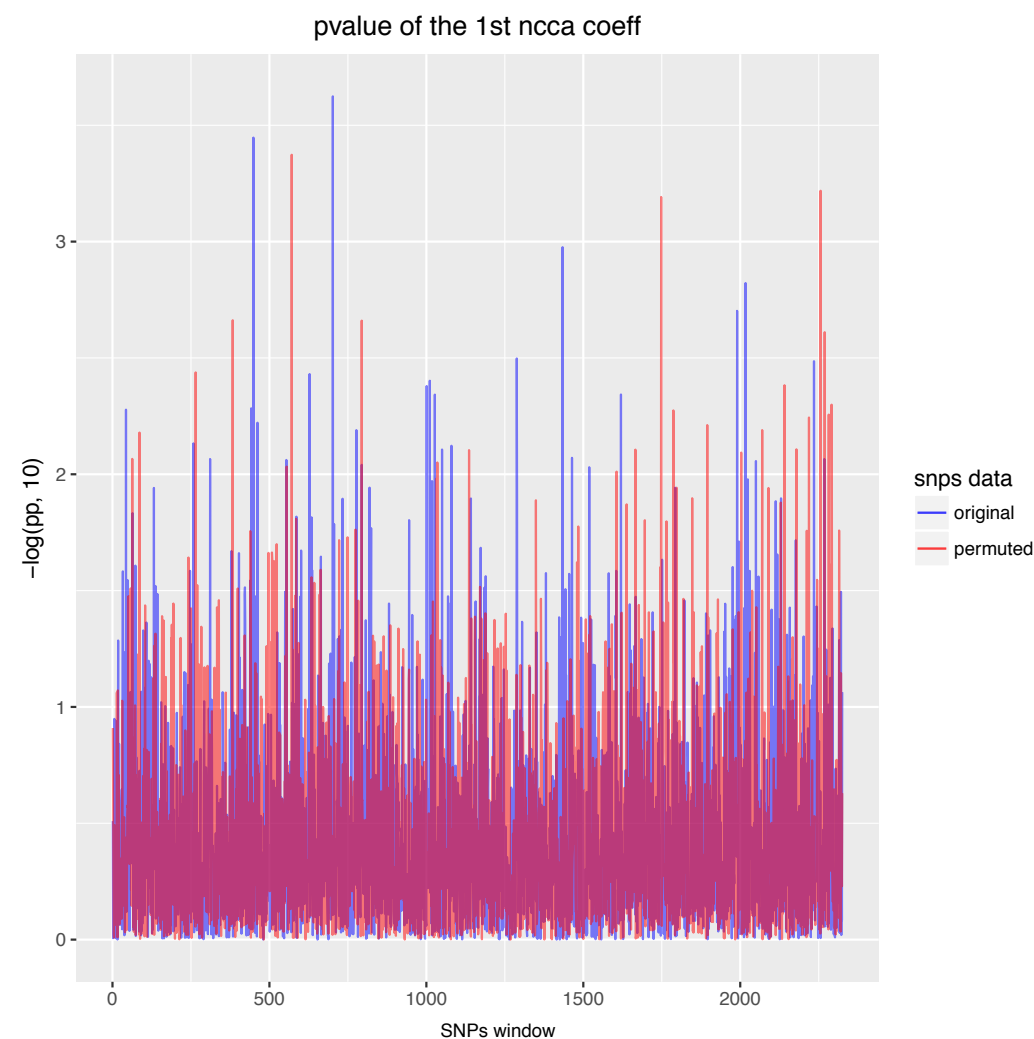
Country	tot.	SNPs
ags	1127	116256
fin	324	164505
hol	232	130827



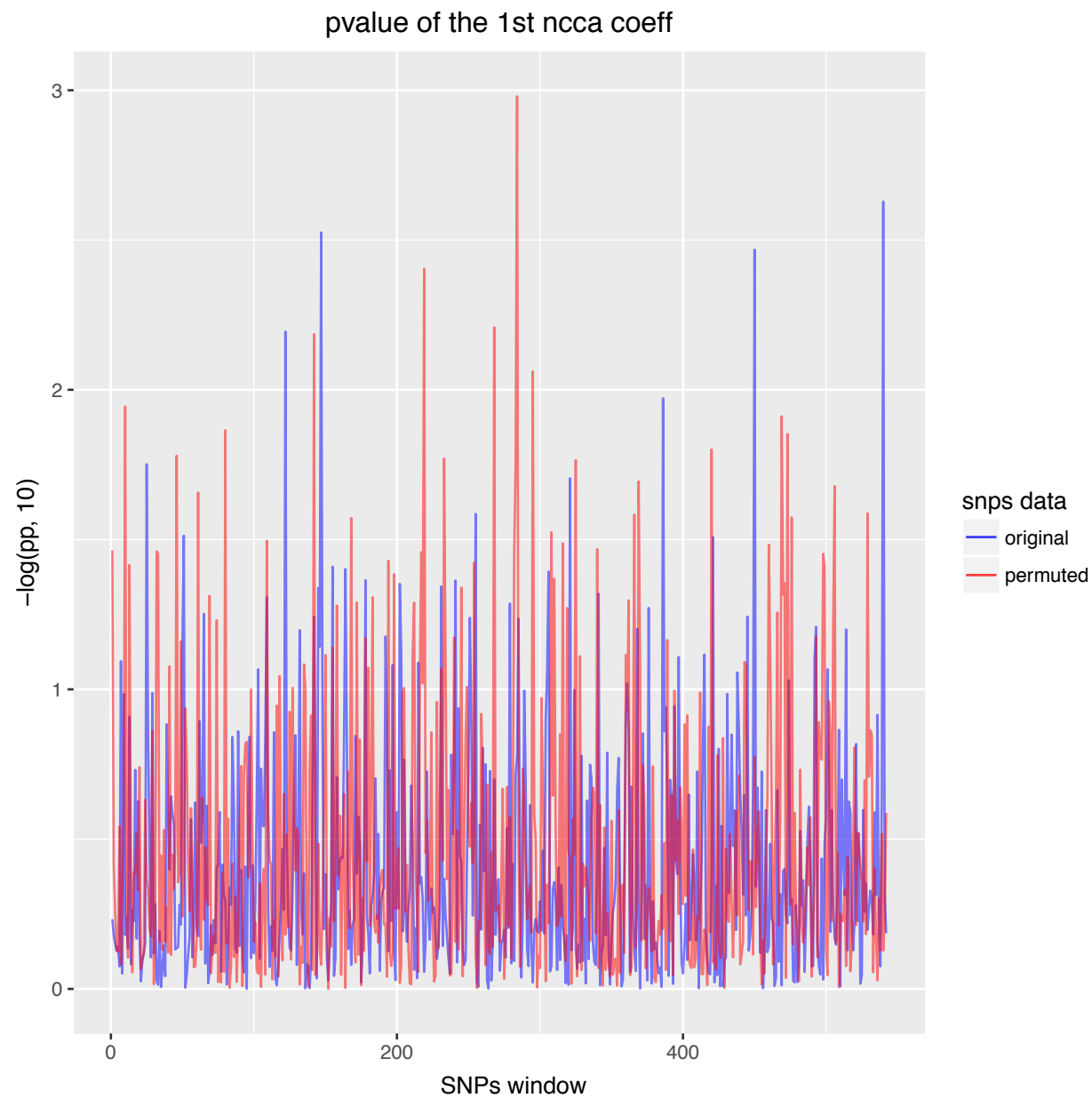
# Fixed sliding window genotype - phenotype analysis



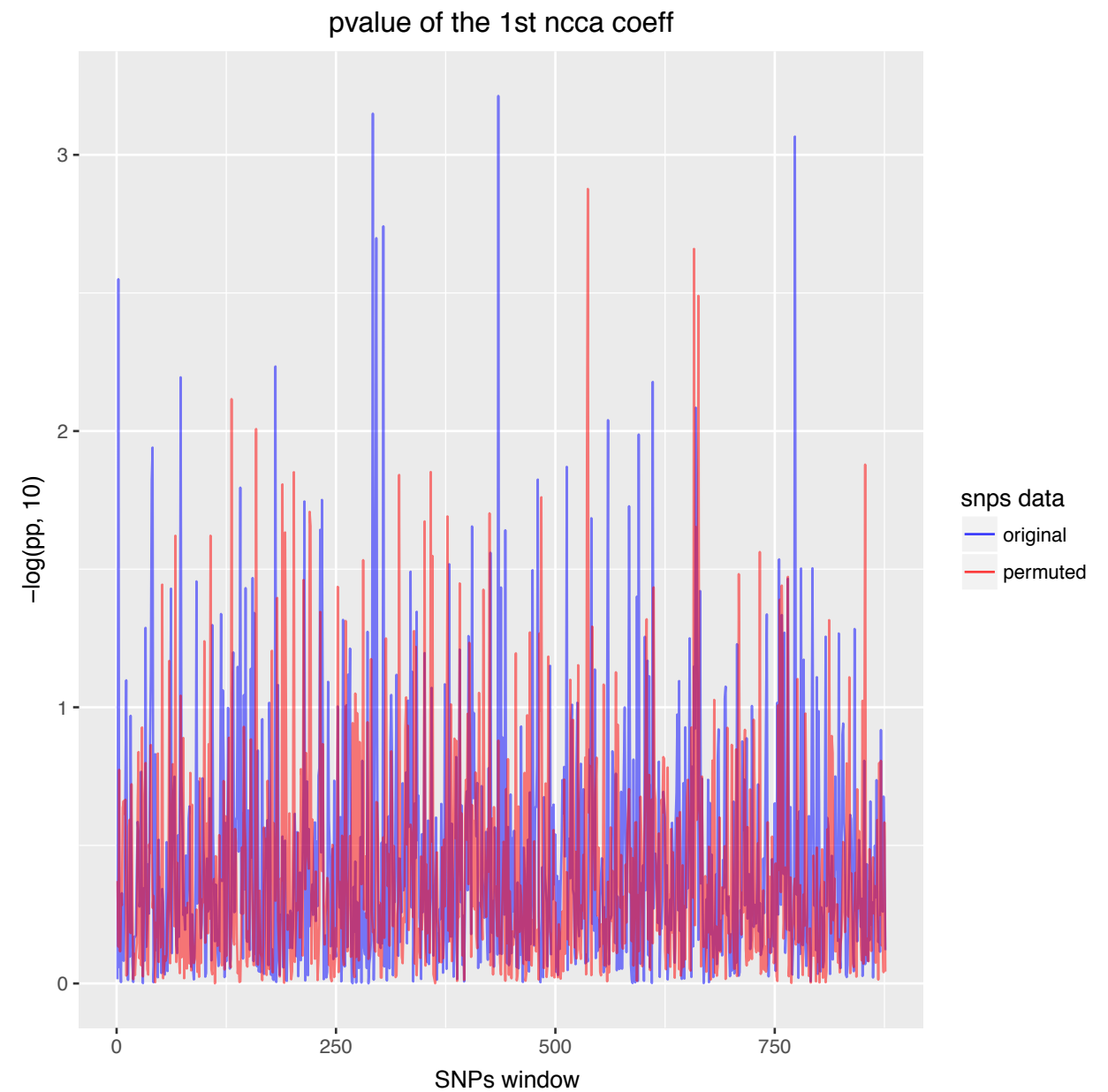
AGS  
70% train, 30% test  
test results



Train the algorithm on AGS data → Apply trained projections on new datasets



Finland



Holland

# Fixed sliding window genotype - phenotype *pairs* analysis

- 100 SNPs vs 2 phenotypes
- 45 phenotypes pairs

smallest p values 1st ncca coefficient pheno pair - SNPs window

pheno pair	AGS	Finland	Holland
1_9	8.94710026517753e-05	5.91008242434299e-05	0.000953834766024619
3_4	0.000154633222918754	3.52949277511172e-05	0.000484424322801713
3_5	3.37191036623441e-05	0.00359810876349331	0.000883104434246107
3_6	0.000633477658925852	4.38931205160031e-05	0.000239652057659967
3_7	0.000413722553299189	0.00531509617684156	0.000276435066801528
3_8	8.89852411845027e-05	0.000842239250596011	0.00080290694893681
5_6	0.000181776179200566	5.18759593954865e-05	6.22549350504098e-05
6_7	0.000386939081757909	0.0046467251046225	7.13167746747567e-05
6_8	8.79188199557346e-05	0.00146918973130294	0.000921721186726954
6_9	0.000775853337516841	0.000168751090801722	0.000437043859029096
7_10	0.000134082259051649	9.10316205422479e-05	1.16902085212089e-06
7_8	9.71884507707799e-05	0.000910602324073362	9.19539739170361e-05
7_9	0.000553138409451481	0.000354164889564412	2.08531109957555e-06
8_10	0.0001857414063966	0.00042691537525319	0.00113091333033422
8_9	0.00058218627497153	0.000488026970689281	0.00323001819848012
9_10	1.55119370752743e-05	0.000919093767100976	2.91006587109788e-05

Kendall rank correlation coefficient

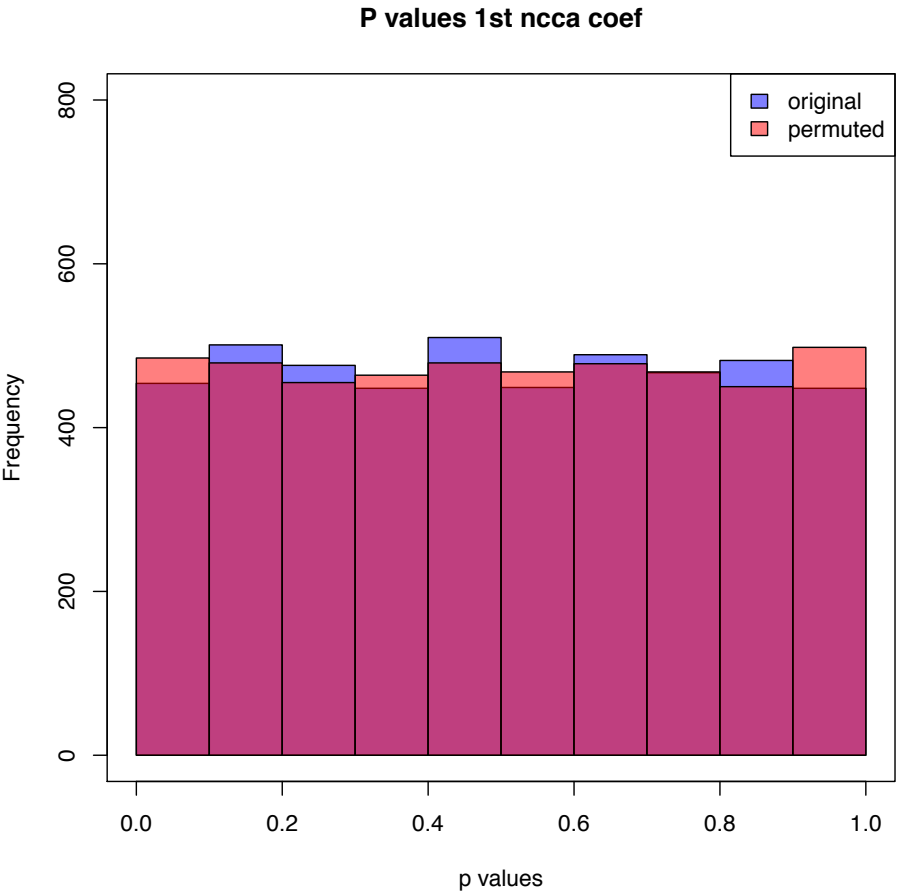
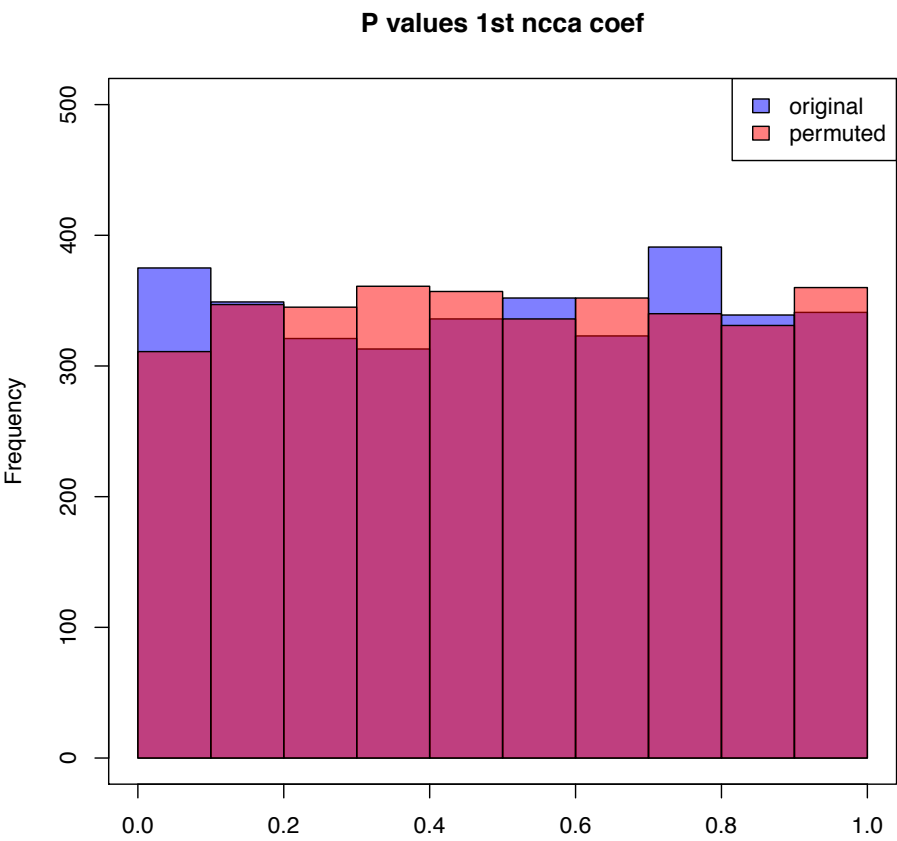
	ags	finland	holland
ags	1	-0.070707070707	0.046464646464
finland	-0.070707070707	1	-0.018181818181
holland	0.046464646464	-0.018181818181	1

# Gene-based approach

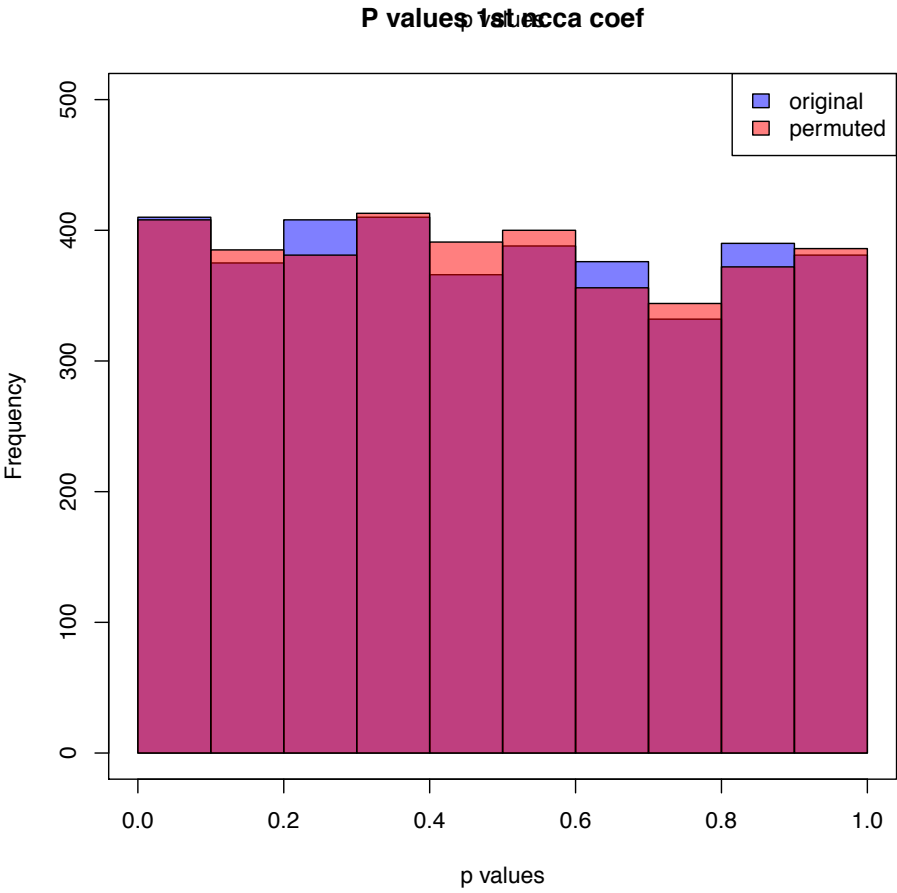
Country	Num of tested genes
AGS	3440
fin	4724
hol	3836

single gene vs10 phenotype

AGS

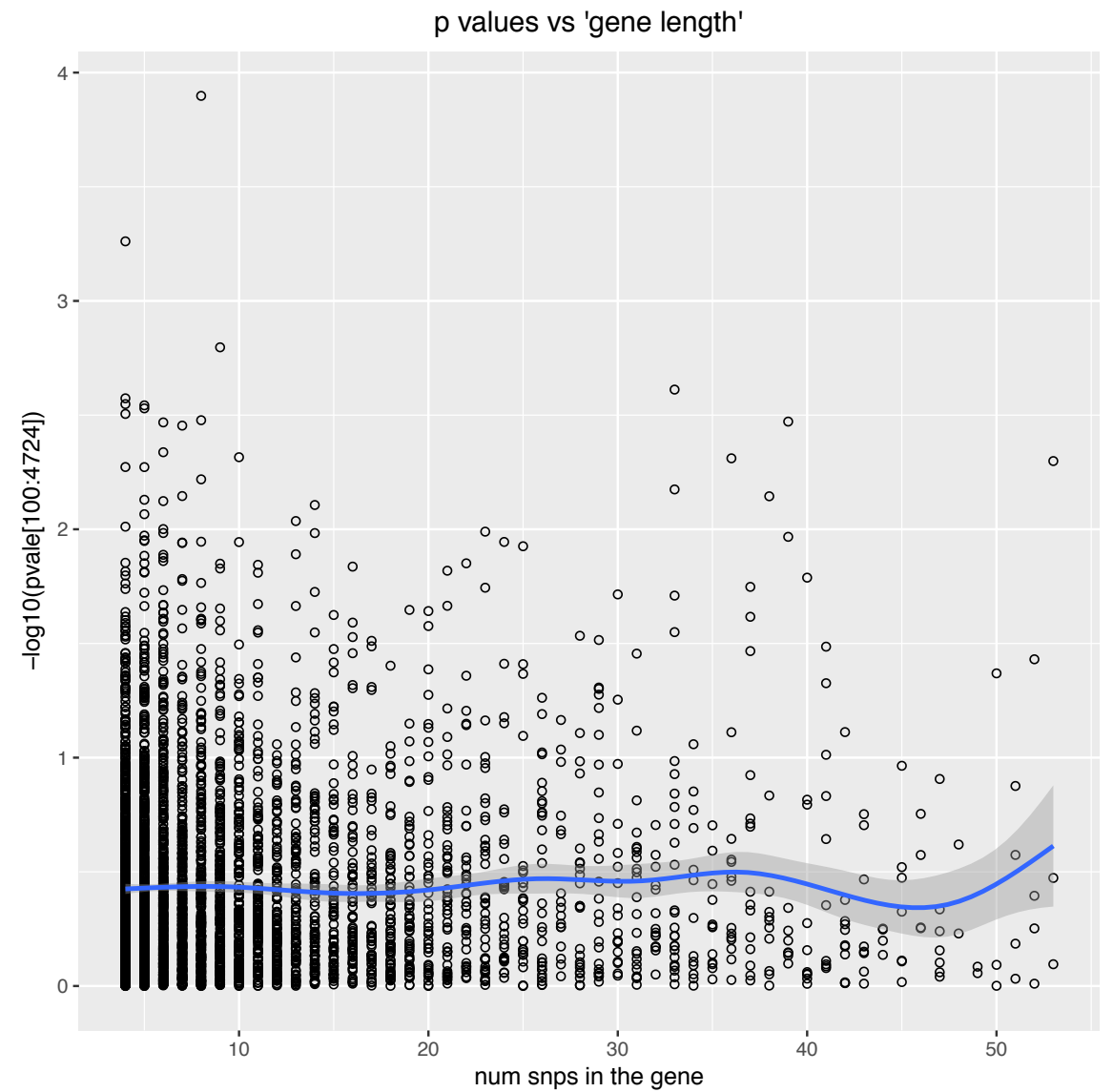
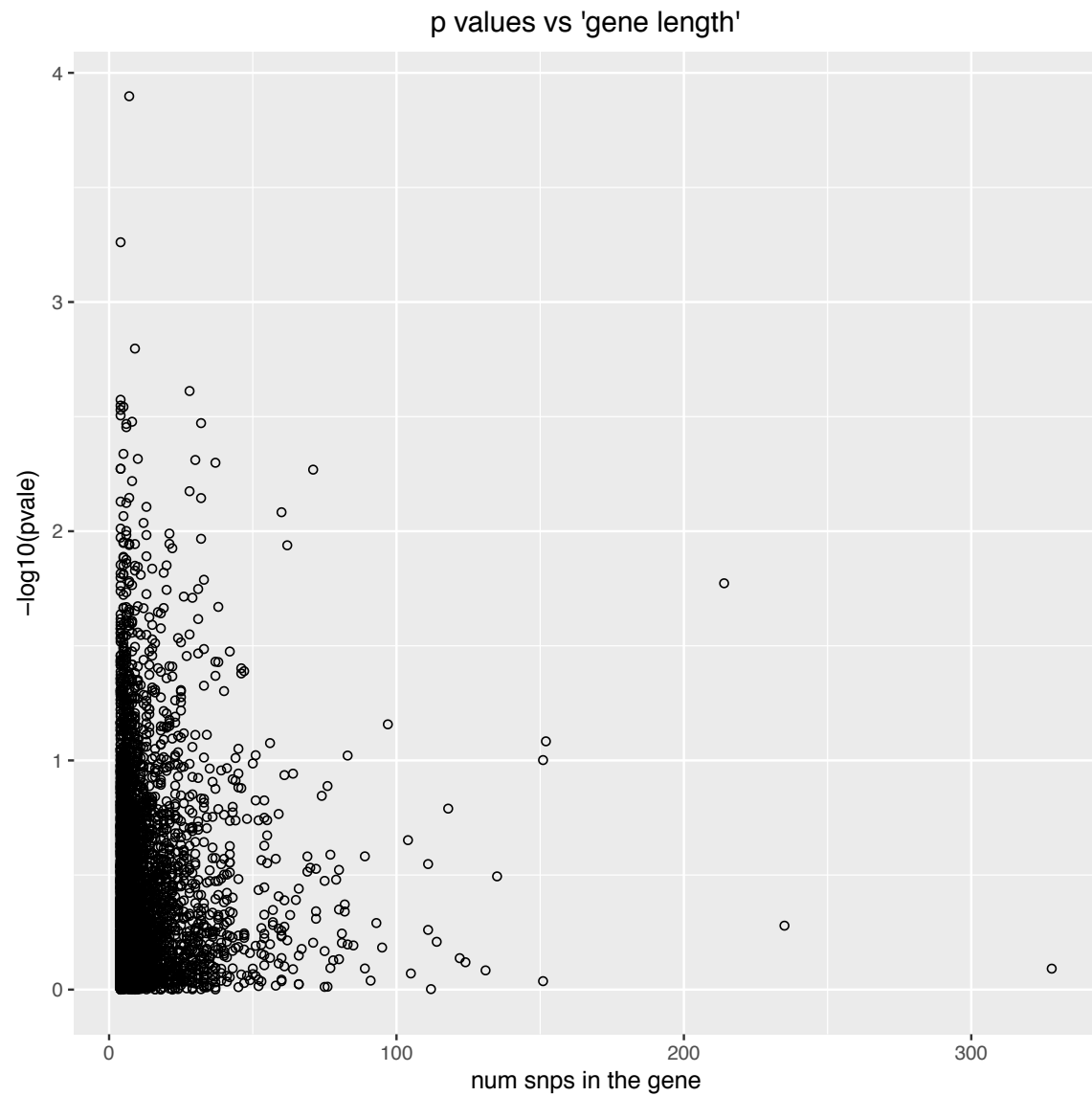


Finland



Holland

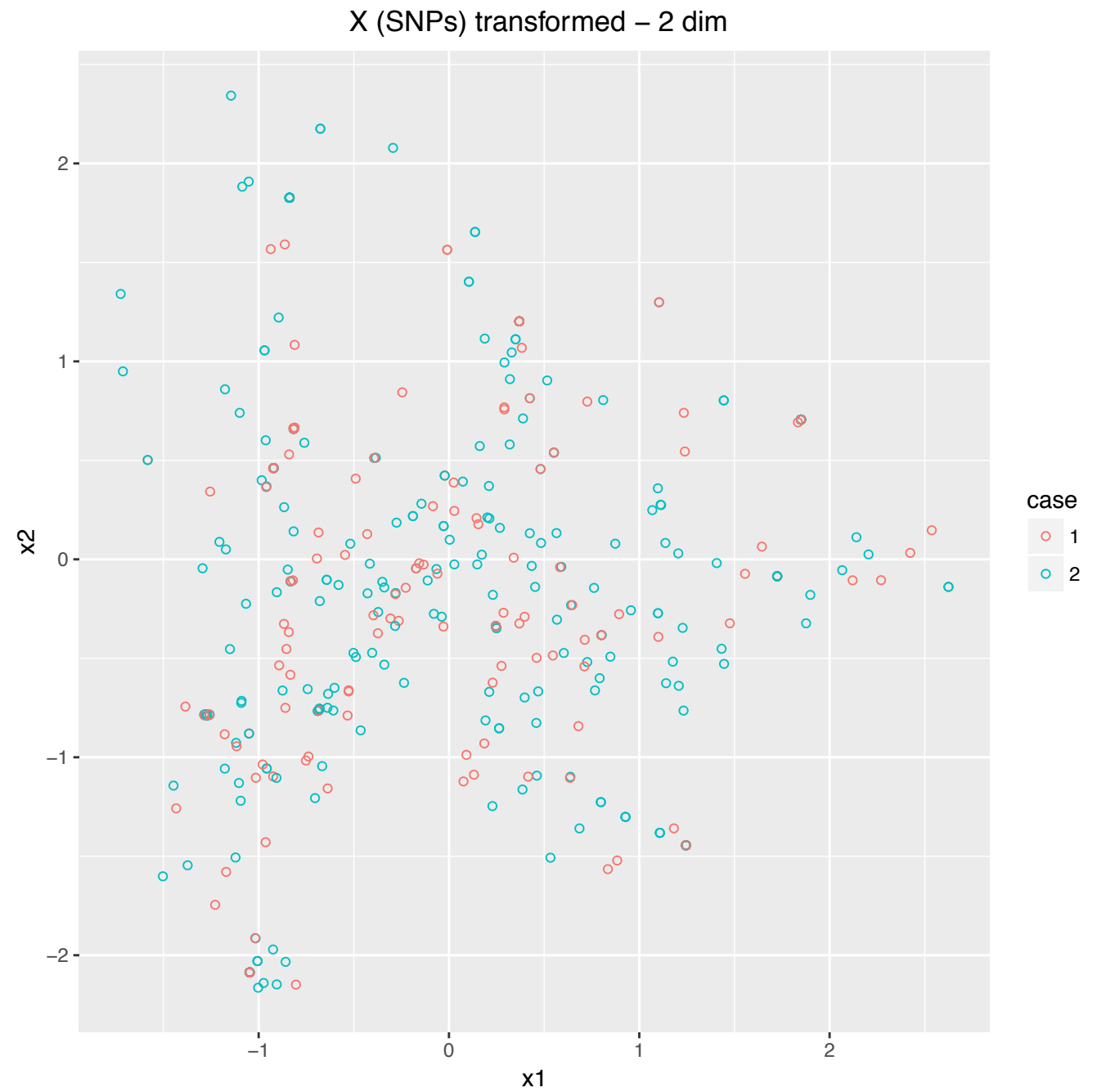
# Relation with number of SNPs in the tested genes



\*Finland cohort

# MLLT4 gene - phenotype association test

AGS cohort  
2 dimensional  
projection of the test  
SNPs dataset



# Future steps

- Identify subsets of SNPs using prior biological knowledge.
- Meta analysis of different cohorts results.
- Code optimisation.
  - Make training and testing independent.
- Cluster analysis of the low-dimensionally projected data.



Thanks!