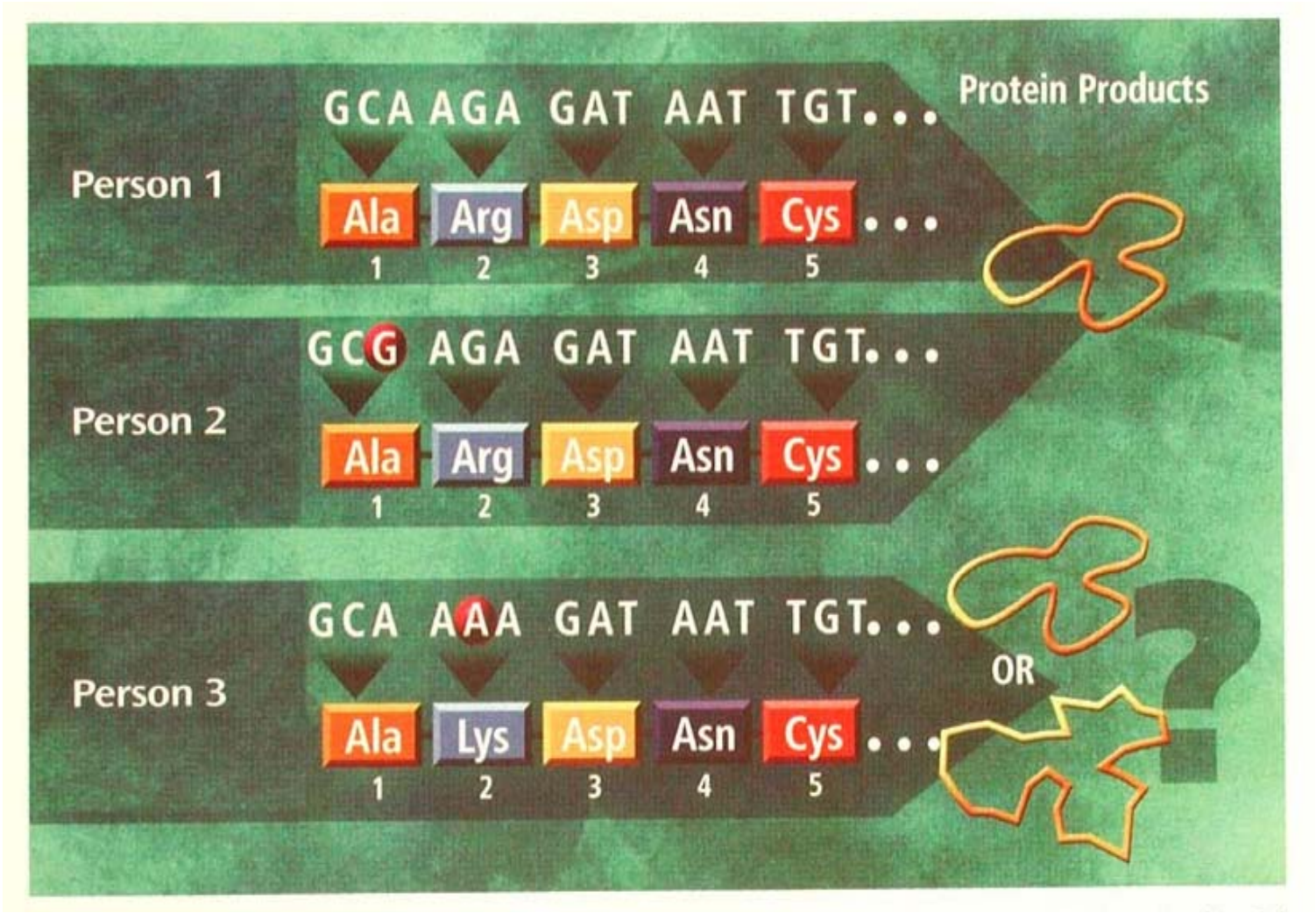
Attention Problems – SNP association

Dorret Boomsma

Toos van Beijsterveldt

Michel Nivard

SNP=single nucleotide polymorphism



Test of association between SNP and trait = test of genetic association

Simple biometrical system:

One SNP with two alleles (di-allelic):

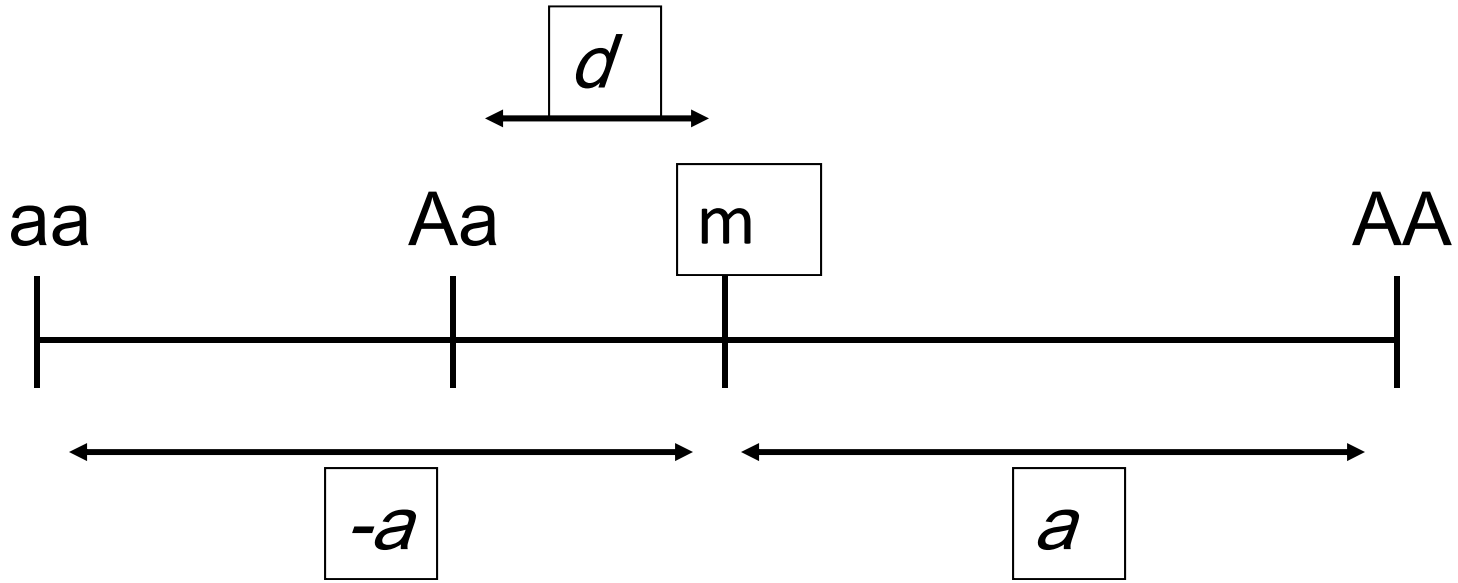
A and a; with frequencies p and q; ($p+q = 1$).

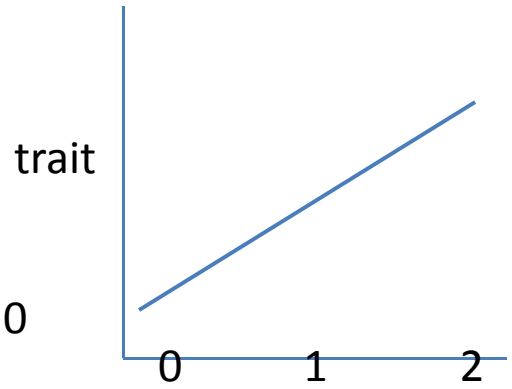
Three genotypes: AA, Aa, aa;

with frequencies p^2, $2pq$ and q^2 ; ($p^2 + 2pq + q^2 = 1$).

AA, Aa and aa are coded as 0, 1, 2 (N of copies of e.g. a)

model

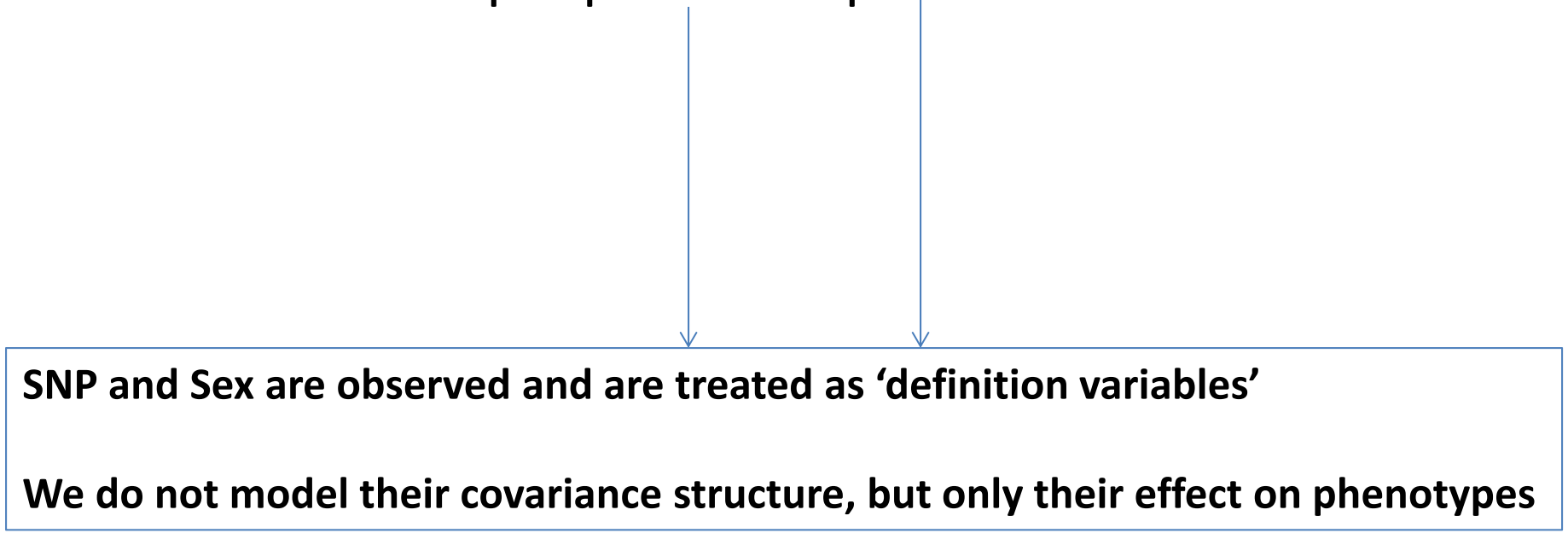




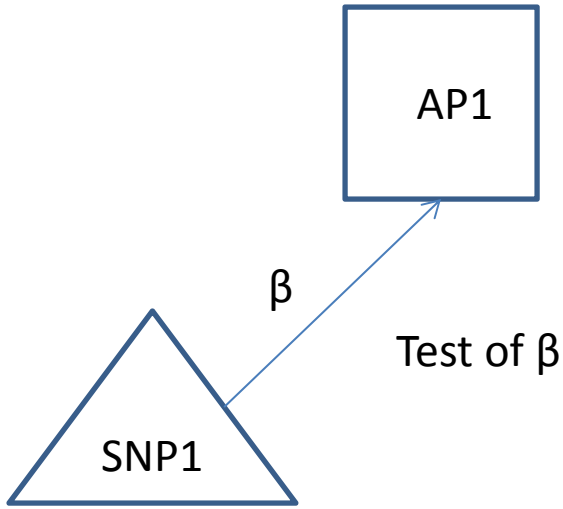
With the 0,1,2 coding we assume that d = 0

$$Y = \text{intercept} + \beta1*SNP + \beta2*Sex + \text{residual}$$

$$ADHD\ score = \mu + \beta1*SNP + \beta2*Sex + \varepsilon$$



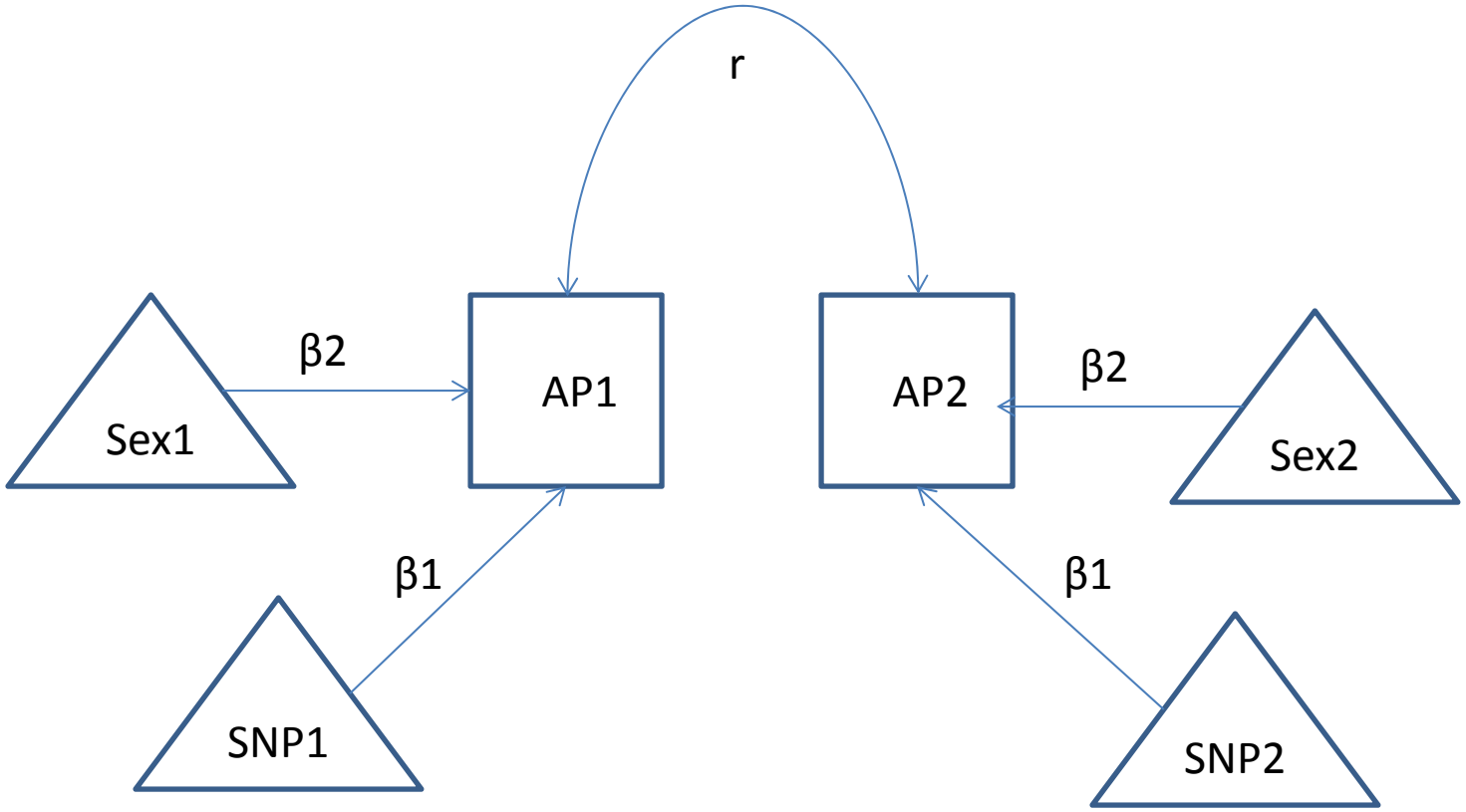
Are Attention Problems influenced by measured genotypes?



Test of β = 0 gives a test of the significance of the SNP

The value of the SNP (0,1,2) MUST be present in the data file for each subject

Attention Problems in twin1 and twin2 are correlated (as a function of zygosity)



Data file

```
rs6265t1 rs4680t1 rs6314t1 rs1007023t1 rs12231356t1 SEXt1 mo3t1 fa3t1 mo7t1 fa7t1 mo10t1 fa10t1 mo12t1 fa12t1 selft
1 1 1 0 0 0 0 1.3 . 2.68 . -1.21 . -1.12 -0.88 5.11 17.61 1 1 0 0 0 0 3.23 . 2.7 . 1.45 . -1.05 -0.88 8.77 17.61
2 0 2 0 1 1 0 3.22 3.2 2.68 1.46 4.72 3.76 3.85 2.85 . -9 0 2 0 1 1 0 4.1 5.72 2.7 2.7 4.73 3.74 3.81 3.81 . -9
3 -9 -9 -9 -9 -9 0 . . . . . 17.51 -9 -9 -9 -9 -9 0 . . . . . 17.51
4 -9 -9 -9 -9 -9 0 . . . . . 17.6 -9 -9 -9 -9 -9 0 . . . . . 17.7
5 0 2 0 0 0 0 4.1 4.93 4.68 4.58 5.55 2.81 2.85 5.9 6.07 17.99 0 2 0 0 0 0 4.1 4.08 4.71 4.54 3.84 3.74 1.55 3.81 2
6 0 2 1 0 0 0 . . . . 2.76 . 1.52 . 4.18 17.63 -9 -9 -9 -9 -9 0 . . . . . 17.64
7 1 1 0 0 0 0 9.71 7.43 . . . . 11.24 8.15 4.18 17.43 1 1 0 0 0 0 9.67 8.22 . . . . 8.82 5.9 4.19 17.43
8 1 2 0 0 0 0 . . . . -1.21 . . . . -9 -9 -9 -9 -9 -9 0 . . . . . -9
9 0 0 0 0 0 0 0 -0.58 2.31 2.68 3.71 1.43 4.64 2.85 2.85 3.2 17.47 0 0 0 0 0 0 -0.52 2.32 1.38 3.69 1.45 2.79 3.81 -0
```

Note:

- missing values for phenotypes are “.”
- missing values for definition variables are -9

(!! Take care that Ss with missing definition variables also have missing phenotypes)

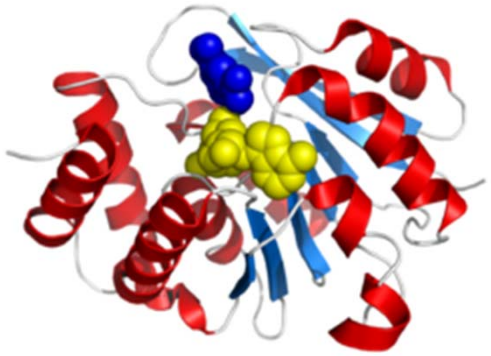
Genotyping in 4 candidate genes:

*mono-aminergic system:
-serotonin receptors (HTR) 2A **(HTR2A)**
-catechol-O-methyltransferase **(COMT)**
-tryptophane hydroxylase type 2 **(TPH2)**

***neurogenesis:
-brain derived neurotrophic factor **(BDNF)**

Genes	SNPs		MAF	p-value HWE
BDNF	rs2049048	A<G	15.6	Ns
	rs7103873	C<G	45.1	Ns
	rs6265	T<C	22.4	Ns
	rs11030107	G<A	24.8	Ns
	rs11030123	A<G	9.2	Ns
	rs1491851	T<C	40.6	Ns
	rs17309930	A<C	17.6	Ns
	rs7124442	C<T	31.6	Ns
COMT	rs4680	G<A	43.8	Ns
HTR2A	rs6311	T<C	42.0	Ns
	rs6314	A<G	8.6	Ns
	rs6313	A<G	44.1	0.008

COMT: Catechol-O-methyltransferase



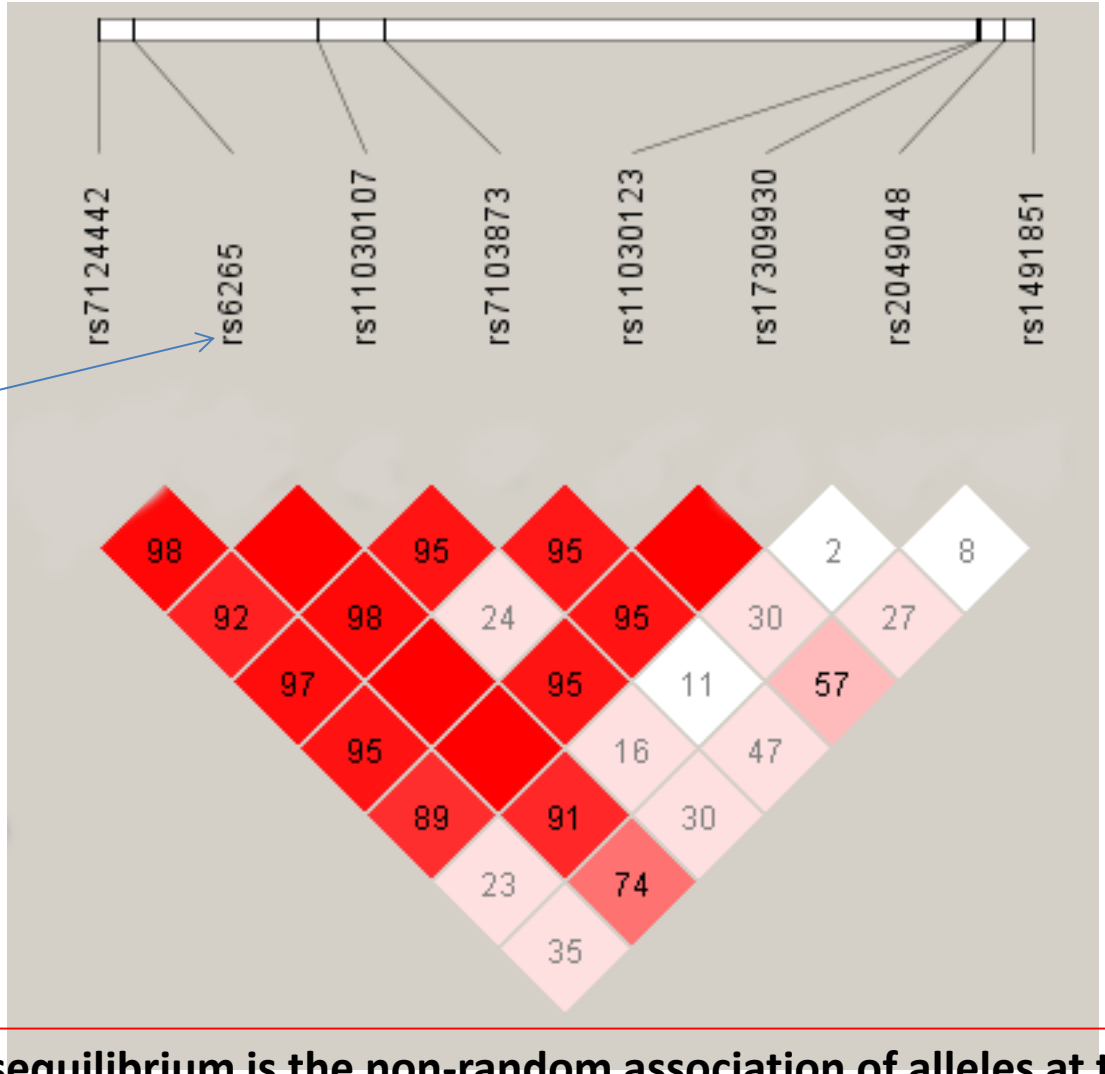
COMT is involved in the inactivation of the catecholamine neurotransmitters (dopamine, epinephrine, and norepinephrine).

A functional single-nucleotide polymorphism (a common normal variant) of COMT results in a valine to methionine mutation at position 158 (Val158Met) [rs4680].

HTR2A 5-hydroxytryptamine (serotonin) receptor 2A

This gene encodes one of the receptors for serotonin, a neurotransmitter with many roles

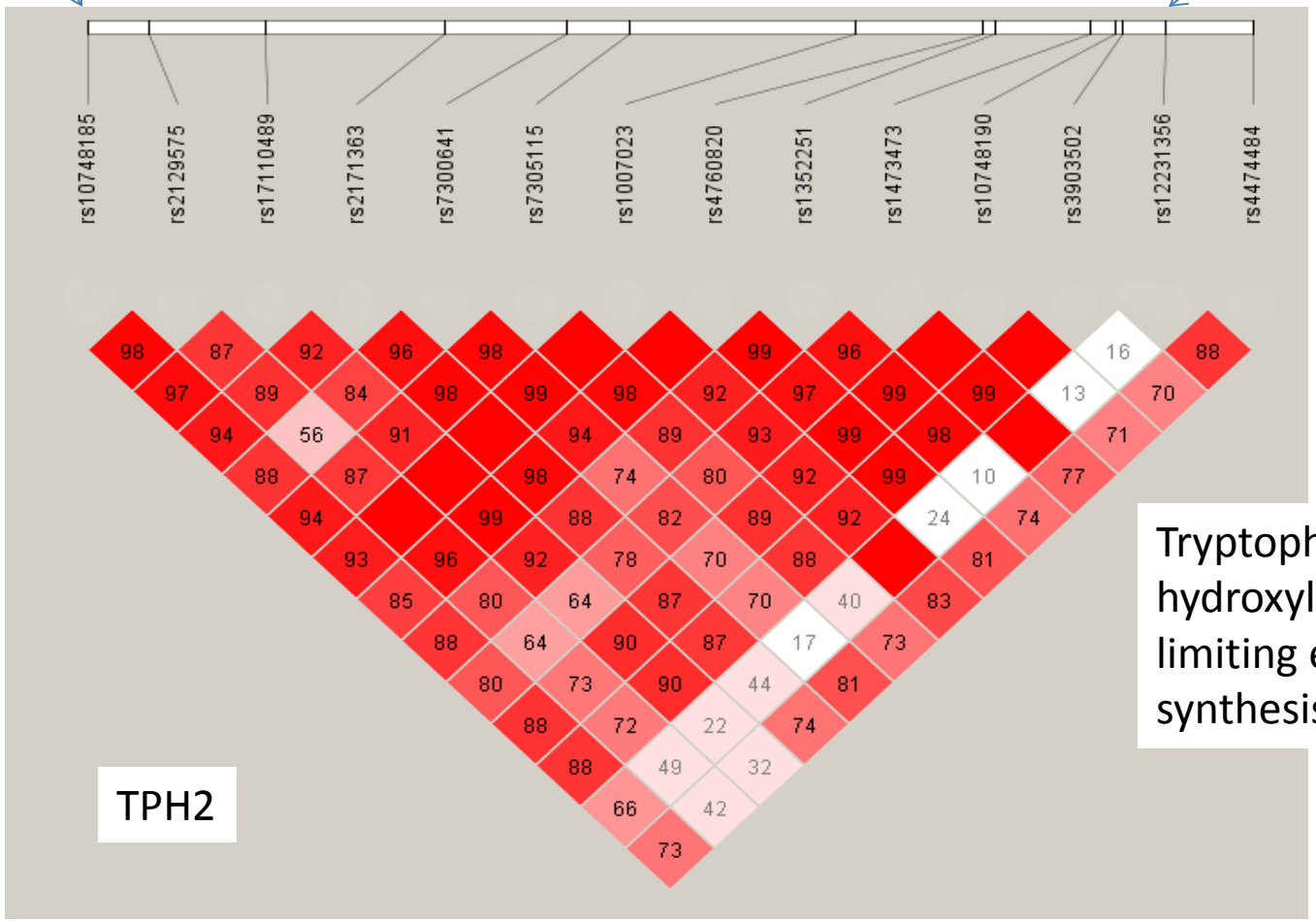
BDNF (Brain-derived neurotrophic factor)
LD plot of D’ between the SNPs



LD: linkage disequilibrium is the non-random association of alleles at two or more loci

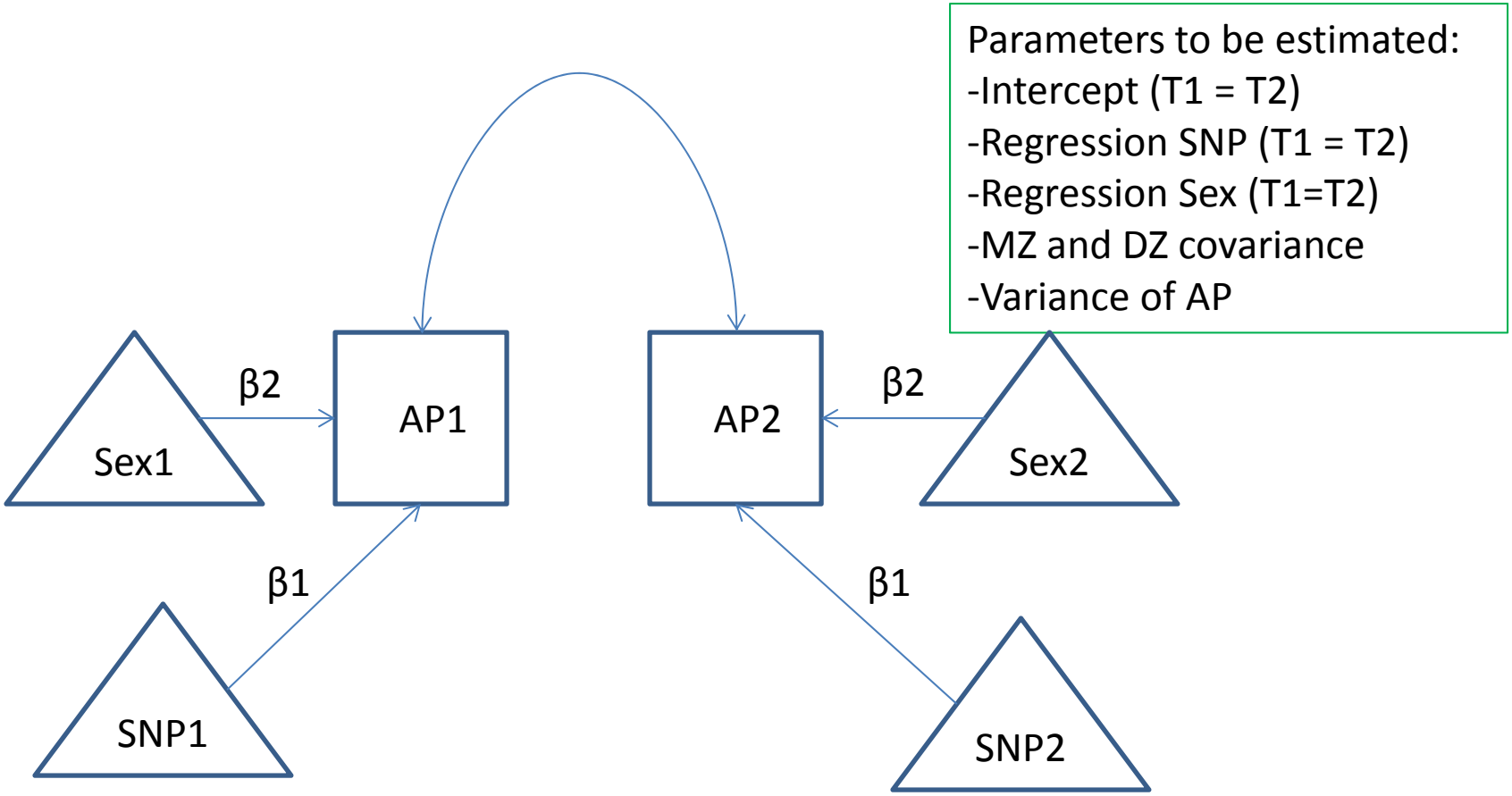
TPH2	rs1007023	G<T	14.9	Ns
	rs10748190	G<A	42.0	Ns
	rs12231356	T<C	7.3	Ns
	rs1352251	C<T	43.2	Ns
	rs1473473	C<T	16.1	0.00005
	rs2129575	T<G	25.4	Ns
	rs2171363	A<G	44.4	Ns
	rs3903502	T<C	41.7	Ns
	rs4474484	A<G	36.9	Ns
	rs4760820	G<C	41.6	Ns
	rs7305115	A<G	44.2	Ns
	rs10748185	A<G	48.7	Ns
	rs17110489	C<T	26.5	Ns
	rs7300641	T<G	17.7	Ns

LD plot of TPH2 indicating D' between the SNPs



Tryptophan hydroxylase is the rate-limiting enzyme in the synthesis of serotonin

Attention Problems in twin1 and twin2 are correlated (as a function of zygosity)



Test of β = 0 gives the significance of the SNP

Exercise

Fit the association model for 1 SNP per run (we consider 5 SNPs, script is for BDNF - SNP).

Try to run for the other 4 SNPs.

Test the significance of the regression coefficients (by comparing constrained and free model).

Univariate means modeling practical

Univariate mean moderation

- Copy the folder univariate means moderation to your drive (its in faculty/Michel)
- Open the R project in R studio.
- It is a five group script. (MZM,DZM,MZF,DZM,DOS)
- First we read in the data, run the script till line 20.
- Then we run “universal” matrices, these are the same in each group (till line 32)

Univariate mean moderation

- On line 41,65,91,116 and 142 we rad in the SNP rs4680 for twin 1 and twin 2.
- In openMX we tell the software to read actual data by adding the prefix ”data.” to the labels.

- Change the snp id into one of the following:
- rs6265, rs4680, rs6314, rs1007023, rs12231356
- Now run the model untill the mxRun command around line: 172.

Univariate mean moderation

- Around line 180 you are required to drop the SNP. Reset the starting value. What should this be according to you guys?

- Also reset the parameter to fixed!

Results

rs6265	
rs4680	
rs6314	
rs1007023	
rs12231356	

Behav Genet (2011) 41:155–164
DOI 10.1007/s10519-010-9406-5

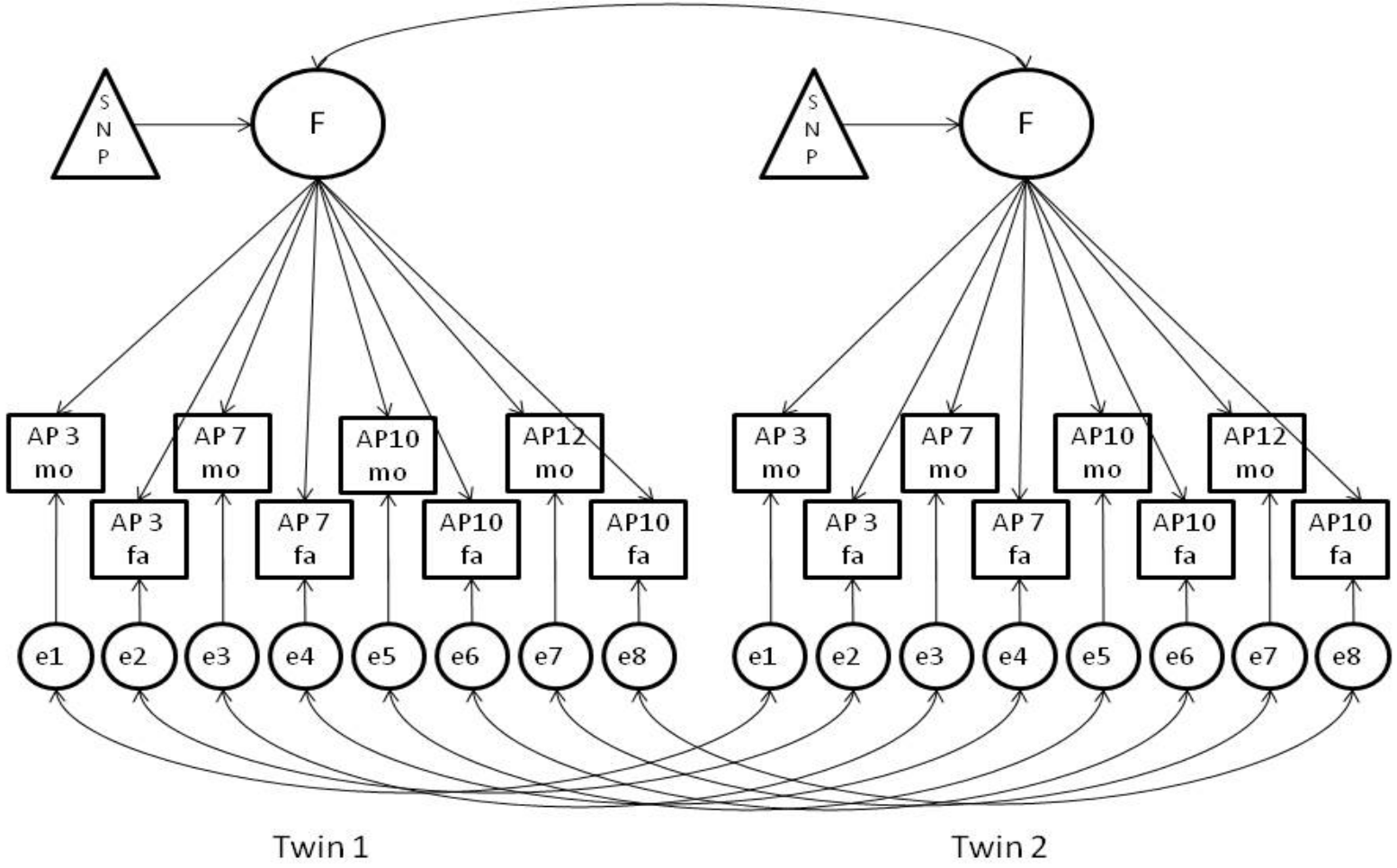
ORIGINAL RESEARCH

Influence of Candidate Genes on Attention Problems in Children: A Longitudinal Study

Catherina E. M. van Beijsterveldt · Christel M. Middeldorp · Margarita C. T. Slof-Op't Landt · Meike Bartels · Jouke-Jan Hottenga · H. Eka D. Suchiman · P. Eline Slagboom · Dorret I. Boomsma

-Genotyping was done in twin pairs (related)
-Phenotypes : ratings by both parents (bivariate)
-Parental ratings at ages 3, 7, 10, 12 (longitudinal)
-Not all children reached at 12 yet (missing data)

Factorial association model for longitudinal Attention Problems in children



Circle = latent (not observed) individual score; square / triangle= observed score; arrow = regression; double headed arrow = correlation

Implementation in Mx

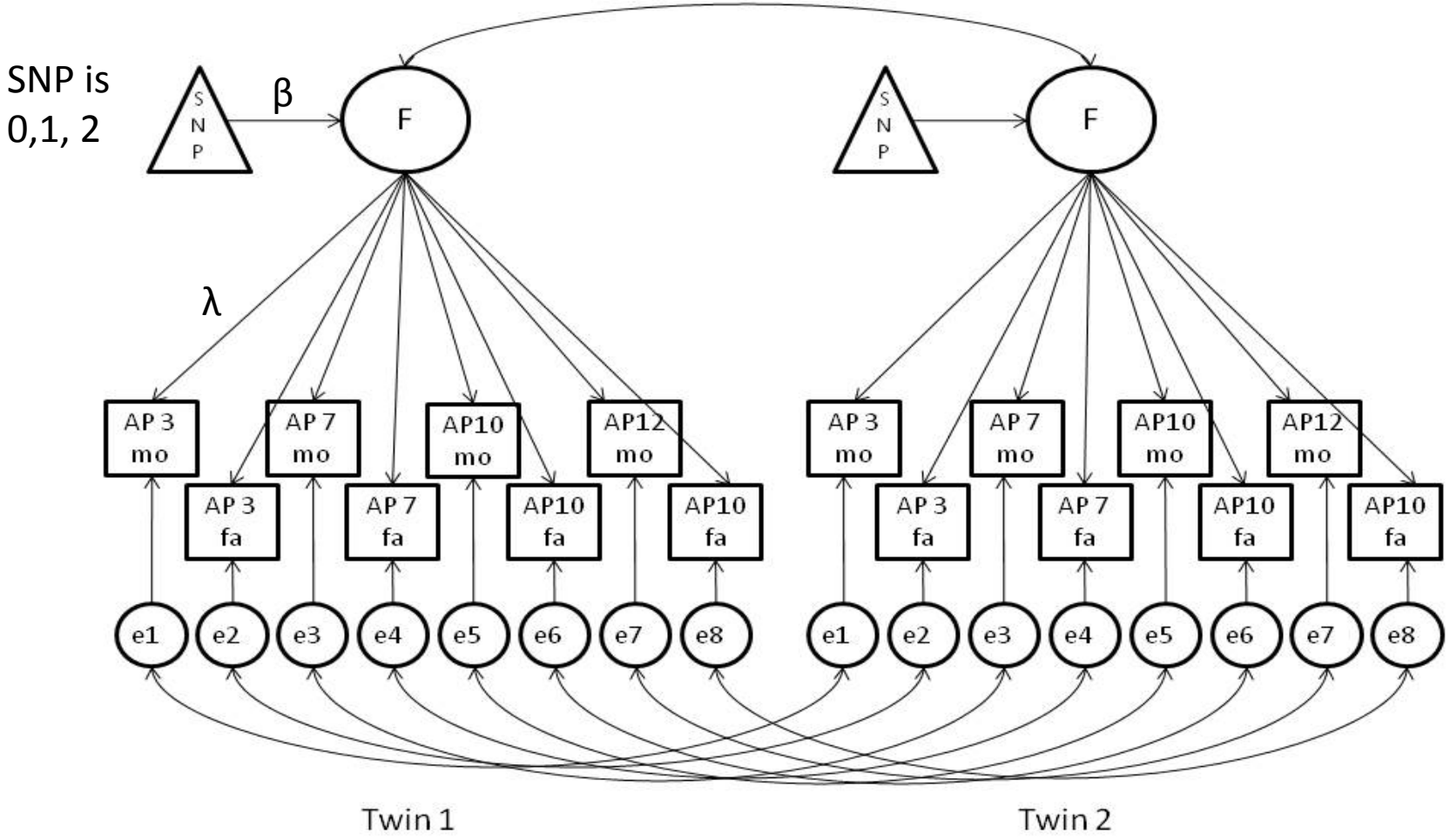
Factor loadings and correlations among latent phenotypes were obtained from running the model in a larger dataset of > 32,000 twins from 16,169 families, who participated at least once:

2,436 MZM, 2,856 DZM,

2,742 MZF, 2,556 DZF, 5,602 (DOS) twin pairs

Age	Rater	Factor Loading	Factor loading Residual	rMZ (residual)	rDZ (residual)
3	Mother	1.2337	1.7753	0.6465	0.1539
	Father	1.2310	1.7186	0.6334	0.1807
7	Mother	2.3359	1.8036	0.5994	0.3254
	Father	2.0936	1.7091	0.6365	0.3872
10	Mother	2.5046	1.7403	0.5652	0.3108
	Father	2.2797	1.6928	0.6110	0.3767
12	Mother	2.3230	1.7810	0.6231	0.3063
	Father	2.1039	1.7563	0.6472	0.4104

Factorial association model : 16 phenotypes (2 twins, 2 raters, 4 time points)



Parameters to be estimated: **effect of SNP,** effect of sex /age /rater, grand mean, twin correlations (for MZ and DZ twins)

Exercise

1 Fit the Factorial association model for 1 SNP per run (consider one of the 5 SNPs).

2 Fit the Factorial association model for all 5 SNPs simultaneously.

The factor model

Use multivariate approaches to model all phenotypic data (all time points / all raters / all indicators) and do not force multivariate data into a single sum score.

Advantage: increase in power

Disadvantage: no standard GWAS software

Increase in statistical power

Ferreira MA, Purcell SM. A multivariate test of association. *Bioinformatics.* 2009, 1;25(1):132-3 (intercorrelations among phenotypes equal)

Medland SE, Neale MC. An integrated phenomic approach to multivariate allelic association. *Eur J Hum Genet*. 2010 18(2):233-9 (factor models)

van der Sluis S, Verhage M, Posthuma D, Dolan CV. Phenotypic complexity, measurement bias, and poor phenotypic resolution contribute to the missing heritability problem in genetic association studies. PLoS One. 2010 ;5(11):e13929 (measurement invariance)

Minica CC, Boomsma DI, van der Sluis S, Dolan CV. Genetic association in multivariate phenotypic data: power in five models. Twin Res Hum Genet. 2010, 13(6):525-43 (also includes longitudinal simplex models)

A = additive genes

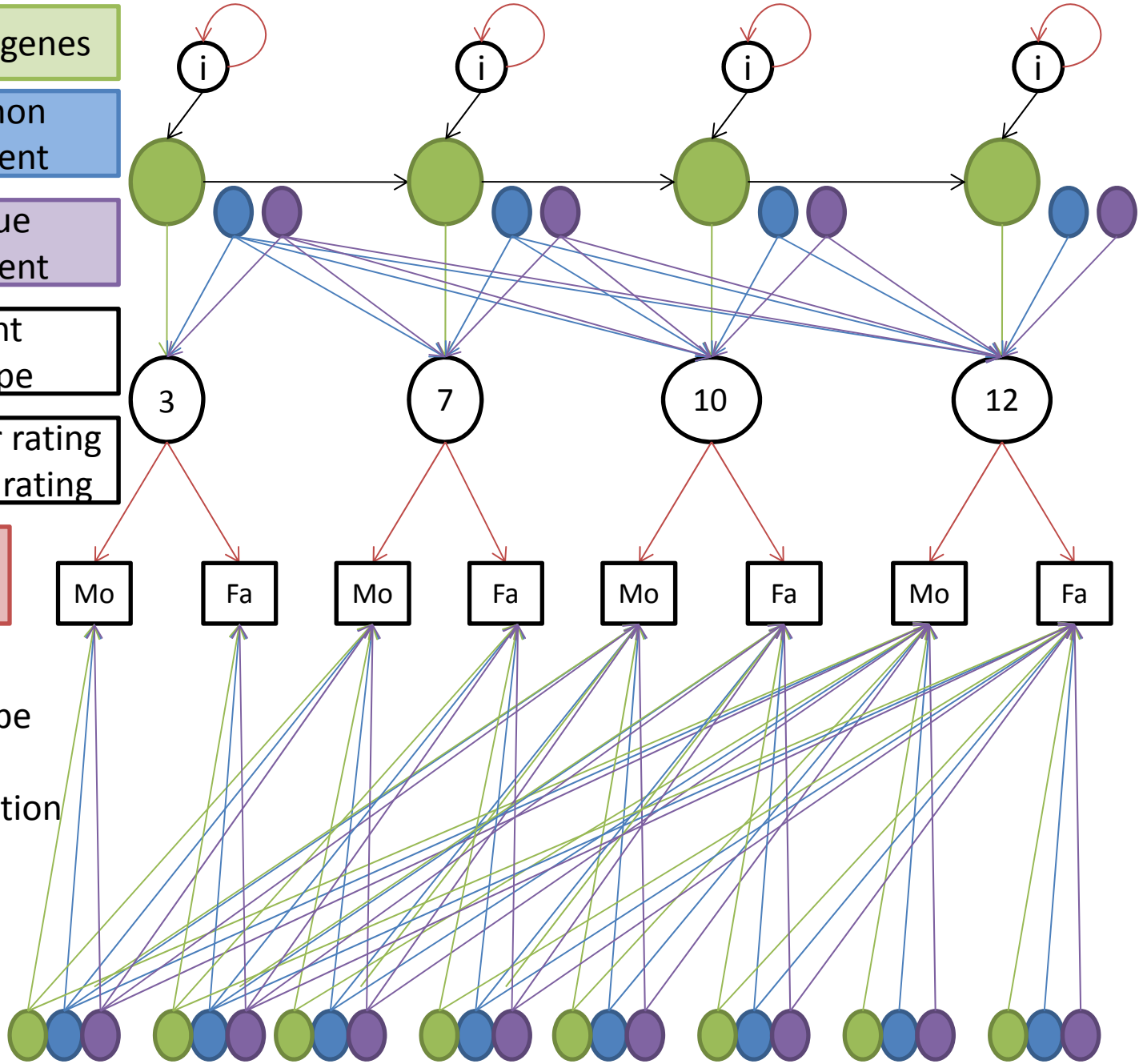
C = common environment

E = unique environment

P = latent phenotype

Mo = mother rating
Fa = mother rating

Parameter in red = 1



Parameters to be estimated:
Loading innovation
Transmission
A,C, E loadings