Regressing SNPs on a latent variable

Michel Nivard & Nick Martin

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)

Genotyping in 4 candidate genes:

*mono-aminergic system:

-serotonin receptors (HTR) 2A **(HTR2A)**

rs6314

-catechol-O-methyltransferase **(COMT)**

rs4680

-tryptophane hydroxylase type 2 **(TPH2)**

rs1007023

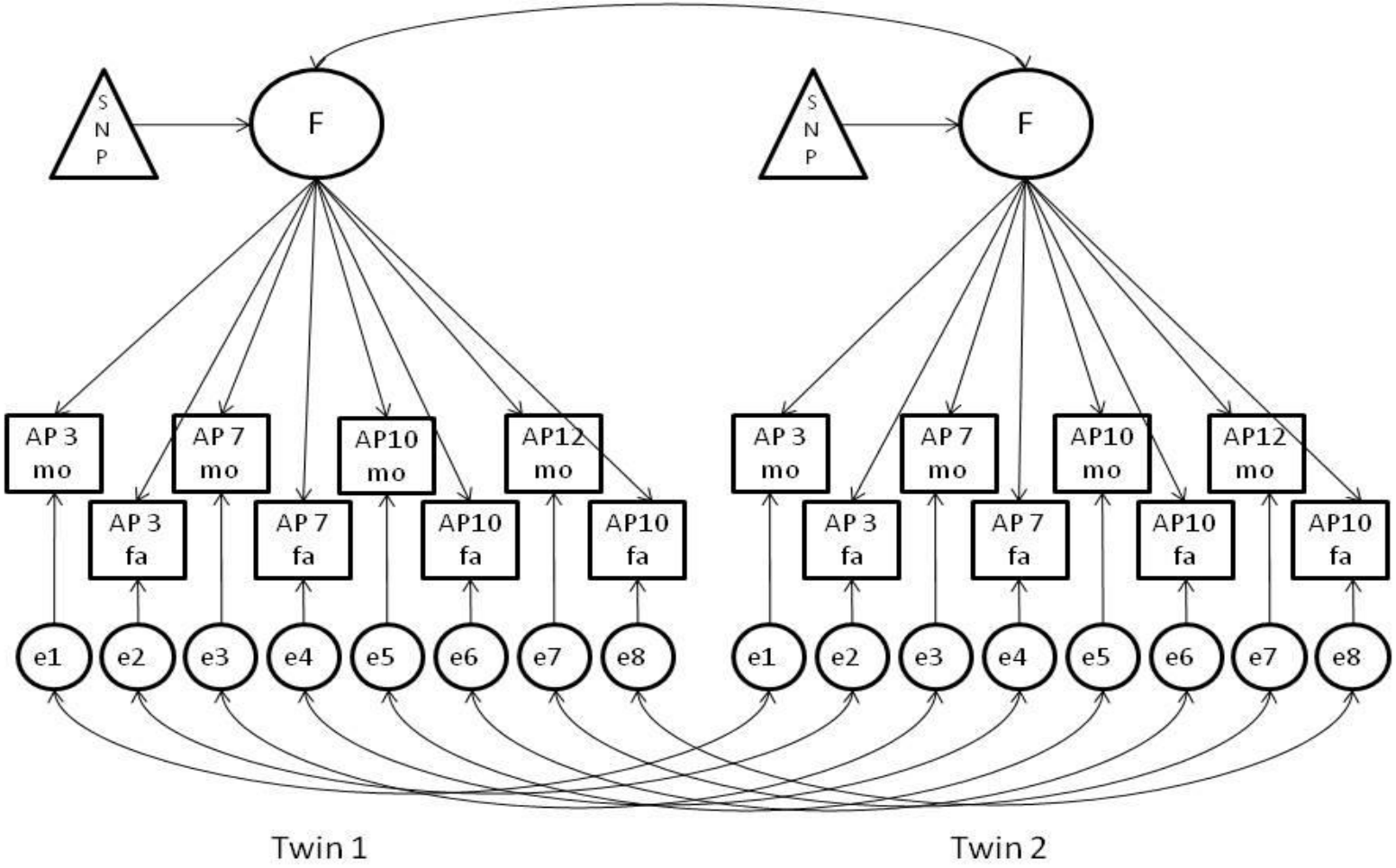
rs12231356

*neurogenesis:

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

rs6265

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

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 (though not always!)

Explicitly model relatedness between subjects

Disadvantage: no standard GWAS software

But why and when should we go for the single factor model and not another model?

Use a single factor model if you believe, or have good reasons to believe, the SNP or gene influences most or all the indicators which load on the factor.

DO NOT Use a single factor model if you believe, or have good reasons to believe, the SNP or gene influences one or only a small number of the indicators load on the factor

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)

Implementation in OpenMx

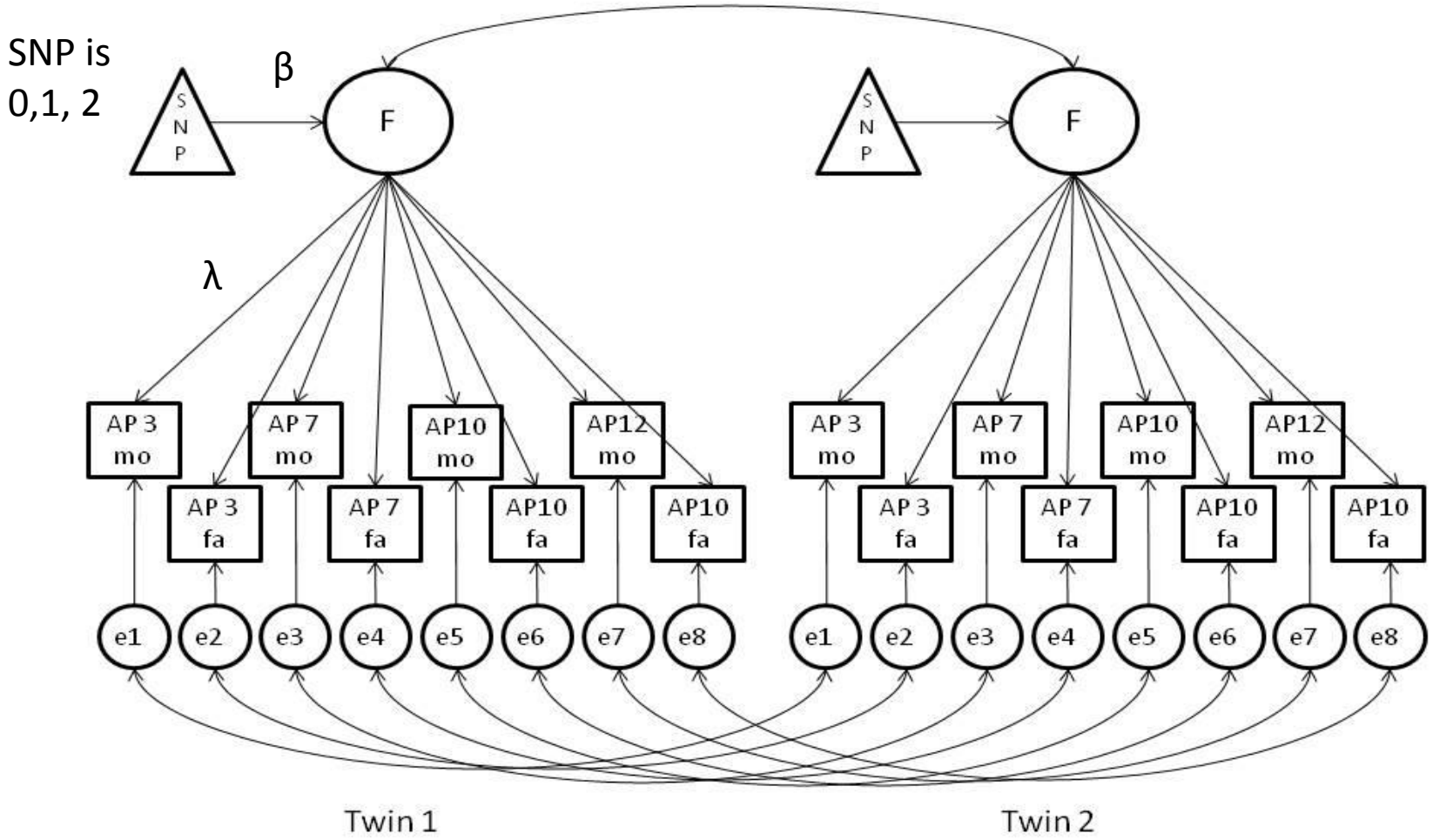
Factor loadings and correlations among latent phenotypes were obtained from running the model in a larger dataset of > 32,000 twins from 16,169 families, who participated at least once:

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.

Fit the model for 1 SNP

- Rs6265 **(BDNF)**
- Rs4680 **(COMT) (Michel)**
- Rs6314 **(HTR2A)**
- rs1007023 **(TPH2)**
- Rs12231356 **(TPH2)**

Fit the model for 1 SNP

```
################################ ASSIGNMENT ################################
# Pick one of the snps out of the dataset and run it in your model: n      #
# below you find there identifiers:                                        #
# rs6265, rs4680, rs6314,rs1007023, rs12231356                              #
################################ ASSIGNMENT ################################


MZSnp <- mxMatrix( type="Full", nrow=1, ncol=2, free=F, labels=c("data.rs4680t1","data.rs4680t2"),name="snp")

MZMean <- mxAlgebra( expression= cbind(All.Mean,All.Mean) + sex%x%All.Beta + (All.BetaSNP%*%snp)%x%t(All.facL), name="mean" )

MZObj <-mxFIMLObjective( covariance="All.covMZ", means="mean", dimnames=manifestVars)

MZModel <- mxModel("MZ",MZData,MZSex,MZSnp,MZMean,MZObj)
```

```
################################ ASSIGNMENT ################################
# Pick one of the snps out of the dataset and run it in your model: n      #
# below you find there identifiers:                                        #
# rs6265, rs4680, rs6314,rs1007023, rs12231356                              #
################################ ASSIGNMENT ################################


DZSnp <- mxMatrix( type="Full",nrow=1, ncol=2, free=F, labels=c("data.rs4680t1","data.rs4680t2"), name="snp")

DZMean <- mxAlgebra( expression= cbind(All.Mean,All.Mean) + sex%x%All.Beta + (All.BetaSNP%*%snp)%x%t(All.facL), name="mean" )

DZObj <- mxFIMLObjective( covariance="All.covDZ", means="mean", dimnames=manifestVars)

DZModel <- mxModel("DZ",DZData,DZSex,DZSnp,DZMean,DZObj )
```

Fit the model for 1 SNP

```
factorNoSNPsModel <- factor1SNPsModel

############ Assignment, create the new values for the BetaSNP matrices ############

factorNoSNPsModel$All@matrices$BetaSNP@free[] = ?????

factorNoSNPsModel$All@matrices$BetaSNP@values[] = ?
```

Results

Rs6265 **(BDNF)**	
Rs4680 **(COMT)**	
Rs6314 **(HTR2A)**	
rs1007023 **(TPH2)**	
Rs12231356 **(TPH2)**	