On the detection of pleiotropic QTLs in extended pedigree data: evaluation of different multi-trait association approaches

Aude Saint-Pierre[(1)], Fabiola Del Greco M[(1)], Peter P Pramstaller[(1,2,3)], Arne Pfeufer[(1)], Cristian Pattaro[(1)]

(1) Institute of Genetic Medicine, European Academy of Bozen/Bolzano (EURAC), Bolzano, Italy - Affiliated Institute of the University of Lübeck, Germany; contact: aude.saintpierre@eurac.edu; (2) Department of Neurology, University of Lübeck, Lübeck, Germany; (3) Department of Neurology, Central Hospital of Bolzano, Bolzano, Italy

Background

- ❑ Joint analysis of correlated phenotypes can provide greater power to map underlying quantitative loci (QTLs) with pleiotropic effects than univariate analysis of the individual phenotypes **[1]**.
- ❑ Various methods to perform multivariate association tests in population- or family-based data have been proposed.
- In GWAS of unrelated individuals, bivariate association analysis based on a **Seemingly Unrelated Regression** (SUR) model **[2]** provides, on average, greater power than univariate analysis **[3]**.
- In data from extended pedigrees, the estimation of the covariance structure makes it difficult to fit bivariate models. An intuitive approach to detect a QTL with pleiotropic effects is to fitting a **univariate model on the first principal component** (PC1) obtained from principal component analysis (PCA) of the phenotypes of interest.

→ Here, we compare different approaches to detect a QTL with pleiotropic effects using the example of two highly correlated cardiac phenotypes measured in an extended-pedigree study.

The SUR model for bivariate association analysis

$$\begin{cases} y_1 = \beta_{10} + g\times\beta_1 + e_1 \\ y_2 = \beta_{20} + g\times\beta_2 + e_2 \end{cases} \rightarrow y = \beta_0 + \begin{pmatrix} g & 0 \\ 0 & g \end{pmatrix} \times \beta + e,\ e \sim N(0, \Sigma \otimes I_n) \text{ where } \Sigma = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}$$

$$\begin{cases} H_0 : \beta_1 = \beta_2 = 0 \\ H_1 : \beta_1 \neq 0 \text{ or } \beta_2 \neq 0 \end{cases} \quad \text{and} \quad F \sim F(2, 2(n-2))$$

Figure 1. QQ plot from bivariate and univariate analysis of traits

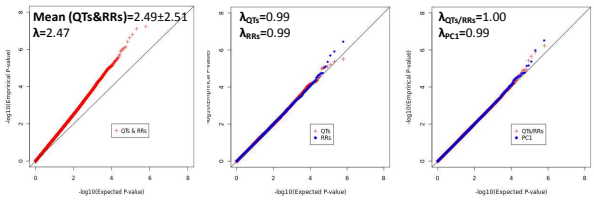

Table 3. Top 11 most associated loci from SUR-based bivariate GWAS: comparison with results from univariate association analyses of QTs, RRs, PC1, and QTs/RRs.

[1]SNP	[2]A1/A2 (Ref. All.)	[3]MAF	SUR	Univariate analysis								[5]Ranks				
				QTs		RRs		PC1		QTs/RRs						
			P	β (SD)	[4]P	β (SD)	[4]P	β (SD)	[4]P	β (SD)	[4]P	SUR	QTs	RRs	PC1	QTs/RRs
Rs1	A/G (G)	0.50	2.4E-07	-0.21 (0.04)	**3.1E-06**	-0.23 (0.04)	**3.5E-07**	0.31 (0.06)	**3.0E-07**	-0.02 (0.03)	0.54	4	1	1	1	166391
Rs2	A/G (A)	0.38	9.3E-07	0.21 (0.05)	8.5E-06	0.19 (0.05)	4.9E-05	-0.29 (0.06)	7.2E-06	0.05 (0.03)	0.08	8	5	19	5	24229
Rs3	A/C (A)	0.09	3.9E-07	0.44 (0.10)	9.7E-06	0.47 (0.10)	**1.2E-06**	-0.65 (0.13)	**1.0E-06**	0.04 (0.06)	0.53	5	7	2	2	162661
Rs4	T/C (T)	0.21	2.8E-06	0.29 (0.06)	4.2E-06	0.22 (0.06)	3.8E-04	-0.36 (0.08)	1.9E-05	0.10 (0.04)	4.7E-03	14	2	124	9	1472
Rs5 (rs2880058, NOS1AP)	G/A (G)	0.32	6.1E-08	0.22 (0.05)	5.9E-06	0.11 (0.05)	2.6E-02	-0.23 (0.07)	3.9E-04	0.13 (0.03)	3.7E-06	1	3	7951	128	4
Rs6	T/C (T)	0.46	8.0E-08	0.01 (0.05)	0.80	0.16 (0.05)	9.6E-04	-0.12 (0.07)	0.06	-0.12 (0.03)	1.4E-05	2	245447	326	19738	8
Rs7	A/G (A)	0.45	1.6E-07	-0.11 (0.05)	2.4E-02	0.02 (0.05)	0.67	0.06 (0.06)	0.34	-0.12 (0.03)	1.2E-05	3	7392	206467	103694	6
Rs8	A/G (A)	0.25	7.4E-07	-0.08 (0.05)	0.13	0.08 (0.05)	0.13	0.00 (0.07)	0.96	-0.15 (0.03)	**1.3E-06**	6	40126	40549	295842	2
Rs9	A/G (A)	0.31	1.3E-06	-0.01 (0.05)	0.79	-0.15 (0.05)	4.2E-03	0.11 (0.07)	0.10	0.11 (0.03)	3.7E-04	11	242977	1313	31506	119
Rs10	C/T (C)	0.32	1.9E-06	0.08 (0.05)	0.11	0.17 (0.05)	4.6E-04	-0.18 (0.07)	7.4E-03	-0.06 (0.03)	3.5E-02	12	35453	156	2324	10755
Rs11	T/G (T)	0.13	8.1E-07	0.05 (0.06)	0.45	-0.10 (0.06)	0.09	0.04 (0.08)	0.60	0.12 (0.03)	2.8E-04	7	137913	27072	184750	97

Genome-wide significant SNPs after a Bonferroni correction are shown in bold; [1]SNP with the lowest bivariate association P-value is reported. [2]A1 is the minor allele; [3]Minor allele frequency; [4]Unadjusted univariate P-values; [5]Ranks of the identified SNP

Materials and Methods

Study sample. In the framework of the MICROS study, that was carried out in 3 isolated villages in South Tyrol (Italian Alps) **[4]**, we consider here 942 individuals with available electrocardiogram (ECG) data. All samples were genotyped with Illumina HumanHap300 SNP microarrays: 306,662 SNPs were available for analysis after quality control (SNP call rate = 98%, individual call rate = 98%, Hardy-Weinberg equilibrium test p-value > 1E-06, minor allele frequency ≥1%).

Phenotypes. The QT and RR intervals, measured by the ECG. Individuals were excluded according to atrial fibrillation and QRS interval >120 ms. QT and RR were standardized before analysis (**QTs** and **RRs**). From QTs and RRs, **PC1** was obtained by means of PCA: PC1= -0.71xQTs-0.71xRRs and explained 90% of the total variance.

Association analysis.

- **Univariate GWAS** on **QTs**, **RRs**, and **PC1** were performed using the *mmscore* function in GenABEL **[5]** and estimating the covariance matrix from the genomic kinship matrix. For biological interpretation we also fit a GWAS on QTs adjusted by RRs (**QTs/RRs**).
- **Bivariate GWAS** on **QTs and RRs** was performed using a SUR model with the Systemfit package, using previously described methods **[3]**, and considering the samples as unrelated.
- All models were adjusted for age, sex, BMI and study location and in all models an additive genetic effect was assumed. The Bonferroni corrected threshold for statistical significance was $3.26\text{x}10^{-6}$.

Results

We report the top 11 bivariate associated loci in **Table 3**. Given the lack of a method to account for the pedigree structure in the bivariate SUR analysis, we provide the ranks of the p-value statistics to allow a fair comparison between methods. In summary:

- ❑ Rs1, 2, and 3 were strongly **associated also with QTs, RRs and PC1**, but not with QTs/RRs. A similar behavior was observed for Rs4 but with a weaker effect on RRs.
  →Simultaneous, strong association of bivariate and both univariate analyses (QTs and RRs) may suggest that an underlying QTL acts on both traits, reflecting a **QTL with pleiotropic effects**.
- ❑ Rs5 was strongly **associated also with QTs, PC1, and QTs/RRs** but not with RRs.
  → Suggestion for an underlying **QTL acting uniquely** on QTs.
- ❑ Rs6, 7 and 8 were also associated with QTs/RRs but not with QTs, RRs or PC1.
- ❑ Rs9, 10 and 11: the association signals were much stronger in the bivariate than in univariate analyses. However, since we could not correct the bivariate analysis p-values for the population structure, the p-value comparison across methods could have limited value.

Discussion/Conclusions

Our bivariate genome-wide association scan using the SUR model appears to be a valuable and powerful screening approach to identify loci with a potential effect on both the traits involved in the analysis.

In fact, four of top 11 loci identified had relevant association statistics also when the individual traits were analyzed separately. The analysis of the first principal component further supported these findings.

In other cases, the signals in the bivariate analysis were driven by one of the two traits, as it was the case for the rs5 located 25 kb upstream of *NOS1AP,* that has been previously identified for the QT interval **[6]**.

Limitation of the SUR method is that the hypothesis testing framework allows for multiple alternatives, i.e.: the null hypothesis can be rejected when the locus is associated with at least one of the two traits. For this reason, the method needs to be coupled with univariate approaches to allow a correct interpretation. Methodological developments to include correction for individuals' relatedness are also ground for further research.

References

[1] Jiang C, Zeng ZB. Genetics **140**:1111–1127 1995
[2] Zellner A . J of the Am Stati Assoc **57**: 348–368, 1962
[3] Saint Pierre A, *et al.* Eur J Hum Genet, 2011 [Epub ahead of print]
[4] Pattaro C, *et al.* BMC Med Genet **8**:29, 2007
[5] Aulchenko YS, *et al.* Bioinformatics **23**:1294-6, 2007
[6] Pfeufer *et al.* Nature Genet, **41**:407–14, 2009