Hierarchical Modelling

Genetic contribution of variants near SORT1 and APOE on LDL cholesterol independent of obesity in children

Gross A, Scholz M

Institute for Medical Informatics, Statistics and Epidemiology

Workshop on "Computational Models in Biology and Medicine"

Study

- Background: dyslipidemia of obese children
- $N = 594$ individuals
  - Blood lipid phenotypes:

    High density lipoprotein cholesterol (HDL-C)
    Low density lipoprotein cholesterol (LDL-C)
    Total cholesterol (TC)
    Triglyceride (TG)
  - Candidate SNPs:

    rs599839 (SORT1)
    rs3846663 (HMGCR)
    rs3812316 (MLXIPL)
    rs174570 (FADS2)
    rs4420638 (APOE)
    rs6102059 (MAFB)
  - Covariates: age, BMI SDS and sex

Classical genotype-phenotype analysis

- Linear regression for each lipid phenotype and each SNP

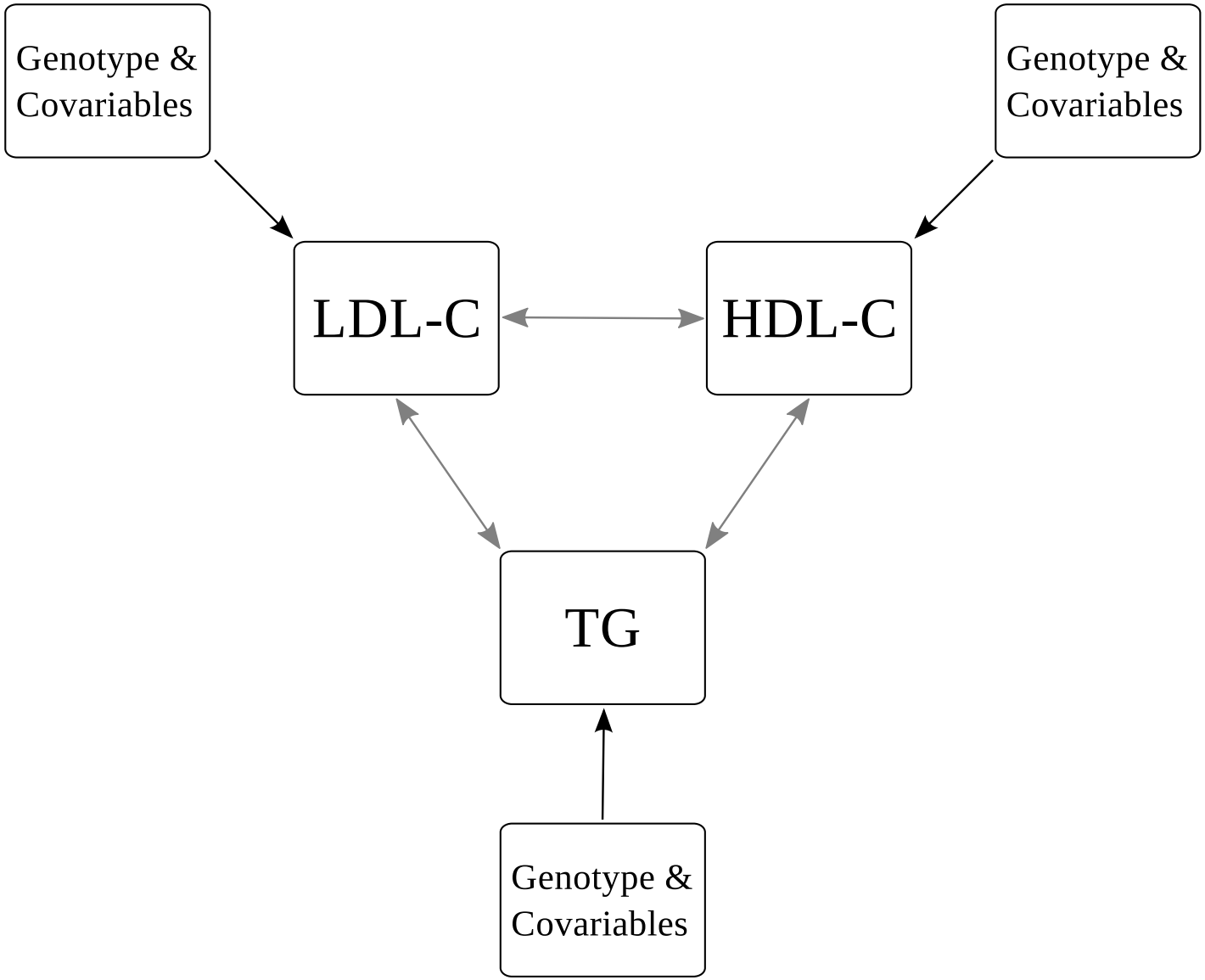
$$
\begin{aligned}
\text{HDL-C} &\sim \text{rs599839} + \text{age} + \text{BMI SDS} + \text{sex} \\
\text{HDL-C} &\sim \text{rs3846663} + \text{age} + \text{BMI SDS} + \text{sex} \\
&\vdots \\
\text{HDL-C} &\sim \text{rs6102059} + \text{age} + \text{BMI SDS} + \text{sex}
\end{aligned}
$$

$$
\begin{aligned}
\text{LDL-C} &\cdots \\
\text{TC} &\cdots \\
\text{TG} &\cdots
\end{aligned}
$$

Phenotype	Variant	N	Beta	SE	p-value
HDL-C	rs599839	576	0.077	0.067	0.25
HDL-C	rs3846663	572	0.098	0.061	0.106
HDL-C	rs3812316	564	0.129	0.093	0.168
HDL-C	rs174570	578	-0.078	0.09	0.387
HDL-C	rs4420638	584	-0.13	0.078	0.098
HDL-C	rs6102059	575	0.038	0.064	0.547
LDL-C	**rs599839**	**576**	**-0.3**	**0.067**	**8.82e-06**
LDL-C	rs3846663	572	0.12	0.062	0.054
LDL-C	rs3812316	564	-0.043	0.094	0.649
LDL-C	rs174570	578	0.021	0.091	0.817
LDL-C	**rs4420638**	**584**	**0.382**	**0.078**	**1.38e-06**
LDL-C	rs6102059	575	0.018	0.065	0.781
TC	**rs599839**	**576**	**-0.257**	**0.067**	**1.50e-04**
TC	rs3846663	572	0.141	0.062	0.024
TC	rs3812316	564	-0.022	0.094	0.812
TC	rs174570	578	-0.04	0.092	0.662
TC	**rs4420638**	**584**	**0.336**	**0.079**	**2.45e-05**
TC	rs6102059	575	0.019	0.065	0.775
TG	rs599839	576	-0.114	0.065	0.081
TG	rs3846663	572	-0.006	0.059	0.913
TG	rs3812316	564	-0.134	0.088	0.127
TG	rs174570	578	0.137	0.087	0.116
TG	rs4420638	584	0.135	0.076	0.076
TG	rs6102059	575	0.075	0.062	0.225

Phenotype	Variant	N	Add Beta	Add SE	Add p-val	Dom Beta	Dom SE	Dom p-val	Rec Beta	Rec SE	Rec p-val
HDL-C	rs599839	576	0.0766	0.0665	0.2499	0.28	0.1696	0.0993	-0.0511	0.0821	0.5339
HDL-C	rs3846663	572	0.0981	0.0607	0.1064	-0.007	0.1241	0.9549	-0.1863	0.0827	0.0247
HDL-C	rs3812316	564	0.1287	0.0932	0.1677	0.3968	0.4029	0.325	-0.1247	0.1004	0.2148
HDL-C	rs174570	578	-0.0775	0.0896	0.3873	-0.0765	0.3483	0.8262	0.0872	0.0983	0.3754
HDL-C	rs4420638	584	-0.1297	0.0782	0.098	-0.4498	0.2746	0.1019	0.1154	0.0872	0.186
HDL-C	rs6102059	575	0.0384	0.0637	0.5474	0.0655	0.1476	0.6574	-0.0427	0.0814	0.6004
LDL-C	**rs599839**	**576**	**-0.2996**	**0.0668**	**8.82e-06**	**-0.4266**	**0.1725**	**0.0137**	**0.3555**	**0.0824**	**1.90e-05**
LDL-C	rs3846663	572	0.1196	0.0619	0.0539	0.3028	0.1262	0.0167	-0.0875	0.0848	0.3025
LDL-C	rs3812316	564	-0.0425	0.0935	0.6494	-0.1414	0.4042	0.7267	0.0406	0.1008	0.6872
LDL-C	rs174570	578	0.0212	0.0913	0.8165	-0.2008	0.3545	0.5714	-0.0415	0.1002	0.6785
LDL-C	**rs4420638**	**584**	**0.3819**	**0.0783**	**1.38e-06**	**0.6728**	**0.2789**	**0.0161**	**-0.4057**	**0.0873**	**4.18e-06**
LDL-C	rs6102059	575	0.018	0.0648	0.7811	0.3394	0.1493	0.0233	0.0739	0.0827	0.3715
TC	**rs599839**	**576**	**-0.2565**	**0.0672**	**1.50e-04**	**-0.2766**	**0.1733**	**0.111**	**0.325**	**0.0828**	**9.67e-05**
TC	rs3846663	572	0.1413	0.0622	0.0235	0.2657	0.127	0.0369	-0.1445	0.0852	0.0904
TC	rs3812316	564	-0.0224	0.094	0.8118	-0.3501	0.4059	0.3888	0.0042	0.1013	0.9667
TC	rs174570	578	-0.0401	0.0917	0.6625	-0.295	0.3561	0.4078	0.0247	0.1007	0.8063
TC	**rs4420638**	**584**	**0.336**	**0.079**	**2.45e-05**	**0.5642**	**0.2805**	**0.0448**	**-0.3597**	**0.0881**	**5.02e-05**
TC	rs6102059	575	0.0185	0.0649	0.7753	0.3292	0.1497	0.0283	0.0699	0.0829	0.3992
TG	rs599839	576	-0.1135	0.065	0.0811	-0.2432	0.1659	0.1433	0.1157	0.0801	0.1493
TG	rs3846663	572	-0.0065	0.0593	0.913	0.0888	0.1209	0.4631	0.0519	0.081	0.5218
TG	rs3812316	564	-0.1342	0.0878	0.1269	-0.8484	0.3784	0.0254	0.103	0.0947	0.277
TG	rs174570	578	0.1367	0.0868	0.1159	0.177	0.3378	0.6005	-0.1504	0.0953	0.1148
TG	rs4420638	584	0.1352	0.076	0.0758	0.3752	0.2671	0.1607	-0.1299	0.0847	0.1258
TG	rs6102059	575	0.0754	0.062	0.2245	0.1639	0.1436	0.2544	-0.0732	0.0793	0.3564

Bayesian hierarchical model



Bayesian model selection

- Multivariate normal distribution of lipid phenotypes

$$(\text{HDL-C}, \text{LDL-C}, \text{TG}) \sim \text{N}_3(\mu, \Sigma)$$

- Phenotype means depend on a variable number of k covariables

$$\begin{aligned} \text{HDL-C} &\sim (\text{rs599839}_{\text{rec}}, \text{rs599839}_{\text{dom}}, \ldots, \text{age}, \text{BMI SDS}, \text{sex})_{k_1} \\ \text{LDL-C} &\sim (\text{rs599839}_{\text{rec}}, \text{rs599839}_{\text{dom}}, \ldots, \text{age}, \text{BMI SDS}, \text{sex})_{k_2} \\ \text{TG} &\sim (\text{rs599839}_{\text{rec}}, \text{rs599839}_{\text{dom}}, \ldots, \text{age}, \text{BMI SDS}, \text{sex})_{k_3} \end{aligned}$$

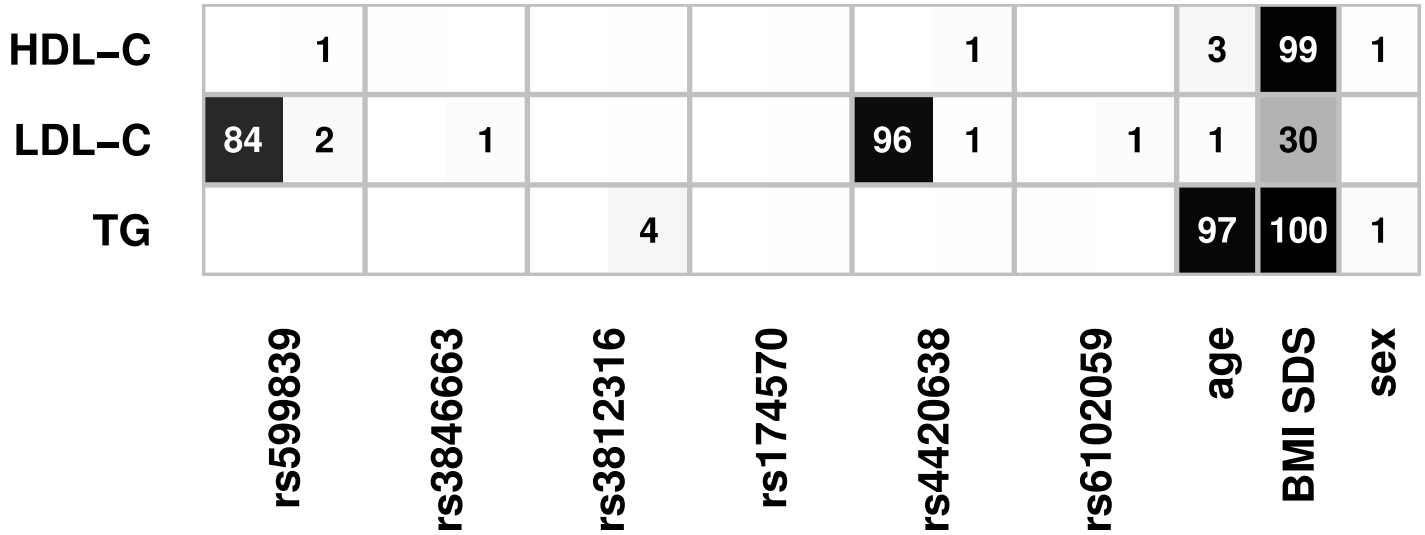
Implementation

- WinBUGS reversible jump

```
01 model {
02
03   for (i in 1:n) {
04     y[i,1:p] ~ dmnorm(y.m[i,1:p],y.t[1:p,1:p])
...
16   }
17
18   for (t in 1:p) {
19
20     for (i in 1:n) { y.m[i,t]<-y.j[t,i] }
21
22     y.j[t,1:n]<-jump.lin.pred(x[1:n,1:c],k[t],b.t)
...
31     k[t] ~ dbin(.5,c)
32   }
33
34   y.t[1:p,1:p] ~ dwish(iii[1:p,1:p],p)
...
49 }
50
51 list(n=594,c=15,p=3,b.t=.0001,...,
52   iii=structure(.Data=c(1,0,0,0,1,0,0,0,1),.Dim=c(3,3)))
```

Phenotype	Model	Probability	Bayes factor
HDL-C	BMI SDS	91.89	371265
HDL-C	age, BMI SDS	3.08	1041
HDL-C	$rs4420638_{dom}$, BMI SDS	0.99	329
LDL-C	$rs599839_{rec}$, $rs4420638_{rec}$	53.49	37691
LDL-C	$rs599839_{rec}$, $rs4420638_{rec}$, BMI SDS	22.88	9720
LDL-C	$rs4420638_{rec}$	7.65	2714
LDL-C	$rs4420638_{rec}$, BMI SDS	4.62	1586
LDL-C	$rs599839_{rec}$	2.54	855
LDL-C	$rs599839_{dom}$, $rs4420638_{rec}$	1.03	340
LDL-C	$rs599839_{rec}$, $rs4420638_{rec}$, age	0.8	266
LDL-C	$rs599839_{rec}$, $rs4420638_{rec}$, $rs6102059_{dom}$	0.77	254
LDL-C	$rs599839_{rec}$, BMI SDS	0.74	244
LDL-C	null	0.56	186
TG	age, BMI SDS	90.47	311171
TG	$rs3812316_{dom}$, age, BMI SDS	3.66	1247
TG	BMI SDS	2.55	856

Marginal Inclusion Probabilities in %



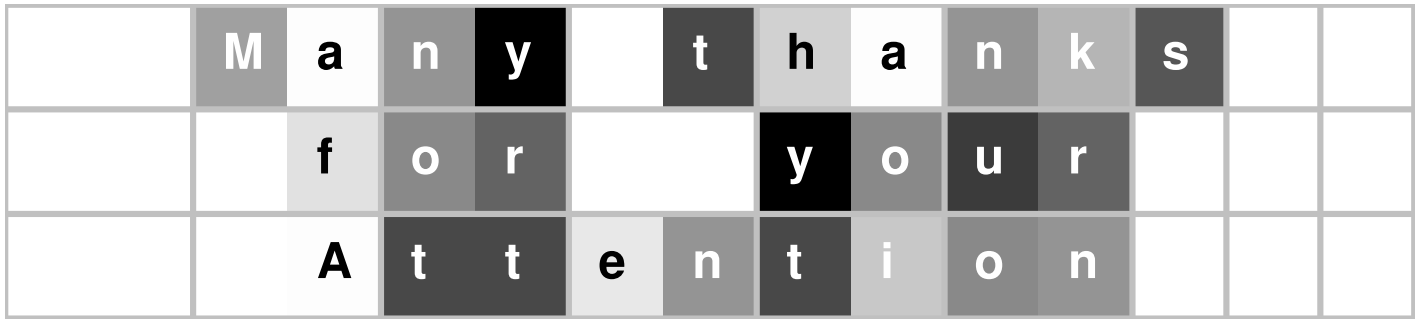
Phenotype	Variant	Probability	Estimate	SD
HDL-C	$rs599839_{dom}$	0.6	0.253	0.178
HDL-C	$rs4420638_{dom}$	1.03	-0.415	0.25
HDL-C	age	3.22	-0.127	0.041
HDL-C	BMI SDS	99.34	-0.21	0.041
HDL-C	sex	0.53	-0.147	0.072
LDL-C	$rs599839_{rec}$	84.3	0.32	0.077
LDL-C	$rs599839_{dom}$	2.2	-0.415	0.171
LDL-C	$rs3846663_{dom}$	1.16	0.258	0.114
LDL-C	$rs4420638_{rec}$	95.57	-0.365	0.081
LDL-C	$rs4420638_{dom}$	0.58	0.347	0.239
LDL-C	$rs6102059_{dom}$	1.23	0.276	0.134
LDL-C	age	1.18	-0.12	0.042
LDL-C	BMI SDS	30	0.146	0.04
TG	$rs3812316_{dom}$	3.81	-0.757	0.346
TG	age	97.28	0.172	0.035
TG	BMI SDS	99.98	0.255	0.044
TG	sex	1.45	-0.166	0.061

Summary

- Strong correlation of rs599839 (SORT1) and rs4420638 (APOE) with blood lipid levels in children
- Advantages of the hierarchical model
  - Accommodation of different phenotypes, genotypes and genetic models
  - Model averaging results in smaller standard deviation of estimates
  - Extensibility
- Bayesian results depend on model structure and prior assumptions

Reference

Genetic contribution of variants near SORT1 and APOE on LDL cholesterol independent of obesity in children. Breitling C, Gross A, Buettner P, Weise S, Schleinitz D, Kiess W, Scholz M, Kovacs P, Koerner A. PLOS ONE. (accepted).


Many thanks
for your
Attention