

BACKGROUND

- **New multivariate method** (MURAT) for analysis of rare-variant associations. It simplifies to SKAT for one phenotype.
- Applied to SBP and DBP in the unrelated data set.
- Compared to SKAT results of separate phenotype analysis.
- $N = 1851$ after removing those subjects missing either phenotype.
- 152,337 exomic SNPs analyzed, only variants with at least 4 copies of the minor allele. Missing genotypes were imputed using MAF.

METHODS

Multivariate Rare-variant Association Test (MURAT)

Suppose for each subject i , we can observe

- K correlated phenotypes, $Y_i = (Y_{i1}, \dots, Y_{iK})^T$;
- Genotype values of a group of v SNPs, $G_i = (G_{i1}, \dots, G_{iv})^T$;
- Covariates X_i .

For region based univariate tests, a single phenotype, Y_{ik} , is linked with G_i and X_i via a linear model $Y_{ik} = \alpha_{0k} + \alpha_k X_i + \beta_k G_i + \epsilon_{ik}$ for $k = 1, \dots, K$. However, for **MURAT**, K phenotypes are associated with genotype and covariates jointly by using a **multivariate linear model**

$$Y_i = \alpha_0 + \alpha(\mathbf{1}_K \otimes X_i) + \beta(\mathbf{1}_K \otimes G_i) + \epsilon_i,$$

where $\alpha_0 = (\alpha_{01}, \dots, \alpha_{0K})^T$, $\alpha = (\alpha_1, \dots, \alpha_K)^T$, $\beta = (\beta_1, \dots, \beta_K)^T$, $\epsilon_i = (\epsilon_{i1}, \dots, \epsilon_{iK})^T$, $\mathbf{1}_K$ is a K -dim vector of ones, and \otimes is Kronecker product. The variant effect, β , in the above multivariate model is assumed to be **normally distributed**,

$$\beta \sim N(\mathbf{0}, \Sigma_\beta),$$

where, the unknown matrix Σ_β is assumed to have a specific form to account for the correlations among Y_{ik} 's. Specifically, it allows a particular **correlation structure** between β_k and $\beta_{k'}$ for $k \neq k'$, such that

- there is a common correlation for the effects of the same variant on different phenotypes, $\text{Corr}(\beta_{kj}, \beta_{k'j}) = \rho$ for variant j .
- the effects of different variants are uncorrelated, $\text{Corr}(\beta_{kj}, \beta_{k'j'}) = \mathbf{0}$ for variant j and variant j'

A score type statistic is derived to test $H_0 : \beta = 0$. With a data-adaptive procedure to manage unknown correlation, ρ , MURAT calculates the p -values analytically, and hence is fast enough to be used for a genome wide analysis.

RESULTS - SINGLE VARIANT TESTS

- All single-variant p -values $< 1e-6$ for SKAT or MURAT are listed in Table 1.
- MURAT appears more powerful.
- rs4926600 is in gene CYP4A22, associated with essential hypertension.
- rs115045946 is near APOC4, associated with coronary artery disease.

Table 1: Top SNPs selected via single variant tests, MURAT and SKAT.

chr	rsID	Nearest Gene	p-value		
			SKAT	DBP	MURAT
1	rs4926600	CYP4A22	3.05e-03	1.39e-01	1.10e-07
1	var_1_91818773		3.83e-01	8.23e-04	8.51e-07
3	rs79314450	HRH1	4.95e-02	4.21e-03	1.75e-08
7	rs116690173	INMT	3.12e-01	2.47e-03	6.23e-07
7	rs79584800	FAM188B	2.91e-01	8.14e-04	2.61e-07
17	var_17_61987931		9.34e-07	6.08e-04	6.26e-06
19	rs138635091	FBN3	1.94e-01	6.33e-04	3.11e-08
19	rs115045946	APOC4	1.09e-01	1.51e-02	5.10e-07

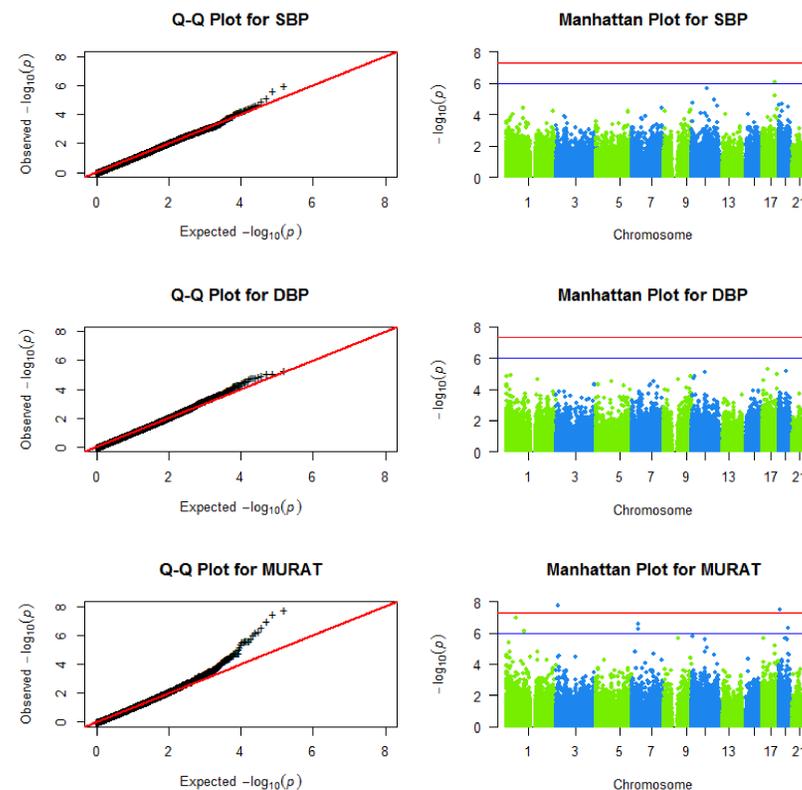


Figure 1: Q-Q plots and Manhattan plots for all individual SNPs

RESULTS - REGION BASED TESTS

Each SNP in Table 1 was grouped with its four nearest neighbours. MURAT again shows improvement over SKAT in SNP sets (Table 2).

Table 2: Results for region based tests when variants in the neighborhood of selected SNPs were grouped together.

Neighborhood of	p-value		
	SKAT	DBP	MURAT
rs4926600	6.82e-02	2.63e-01	1.97e-06
var_1_91818773	7.09e-01	1.30e-03	1.73e-04
rs79314450	2.27e-01	2.82e-02	1.51e-08
rs116690173	7.61e-01	1.39e-02	2.13e-04
rs79584800	4.44e-01	3.91e-03	5.41e-06
var_17_61987931	2.25e-07	8.27e-04	7.13e-05
rs138635091	6.47e-01	5.14e-03	9.90e-06
rs115045946	1.40e-02	5.89e-03	3.77e-03

DISCUSSION

- Multivariate tests are more powerful than univariate tests when multiple phenotypes are correlated.
 - It can lose power (usually mild) when a variant/region is only associated with a subset of multiple phenotypes.
 - eg: In Table 1, the MURAT p -value for var_17_61987931 is larger than the SKAT one for SBP.
- In this study, the region based test does not demonstrate improved power over the single variant test, because it is sensitive to the proportion of causal variants in a region.
 - eg: The results in Table 2 are perhaps expected since there exists only one significant SNP in each set, and region based tests lose power when including neutral variants.

AFFILIATIONS AND ACKNOWLEDGEMENTS

1. Department of Epidemiology, Biostatistics and Occupational Health, McGill University, Montreal, QC, Canada.
2. Lady Davis Research Institute, Jewish General Hospital, Montreal, QC, Canada.
3. Département de Mathématiques, Université du Québec à Montréal, Montreal, QC, Canada.
4. Departments of Human Genetics and Oncology, McGill University, Montreal, QC, Canada. E-mail: celia.greenwood@mcgill.ca

This work was funded by the Canadian Institutes for Health Research (CIHR) MOP 115110.