G x E based on genomic information

S. Hong Lee

(Jan/17)

LINEAR MIXED MODELS AND GENOMIC RELATIONSHIP

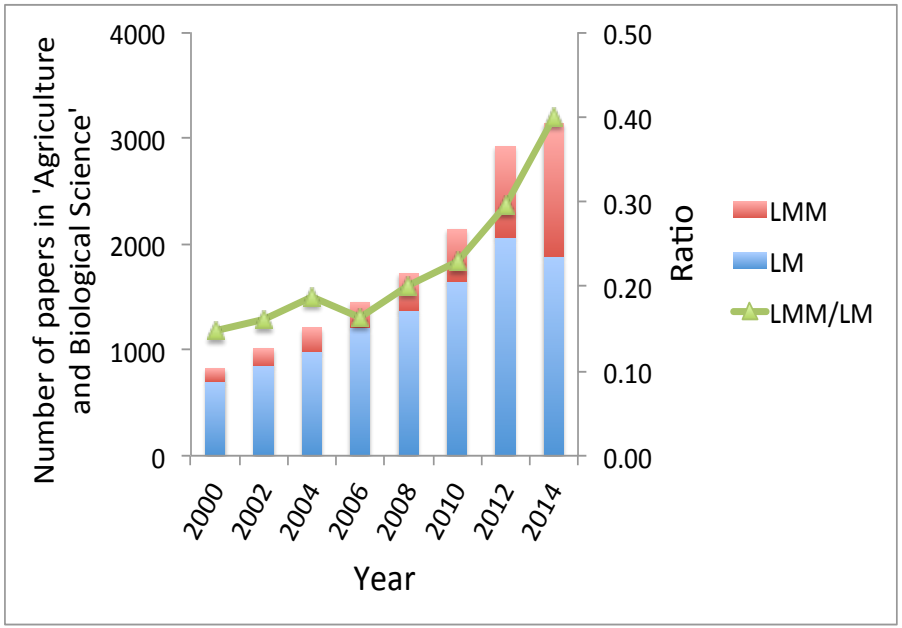
Linear mixed model

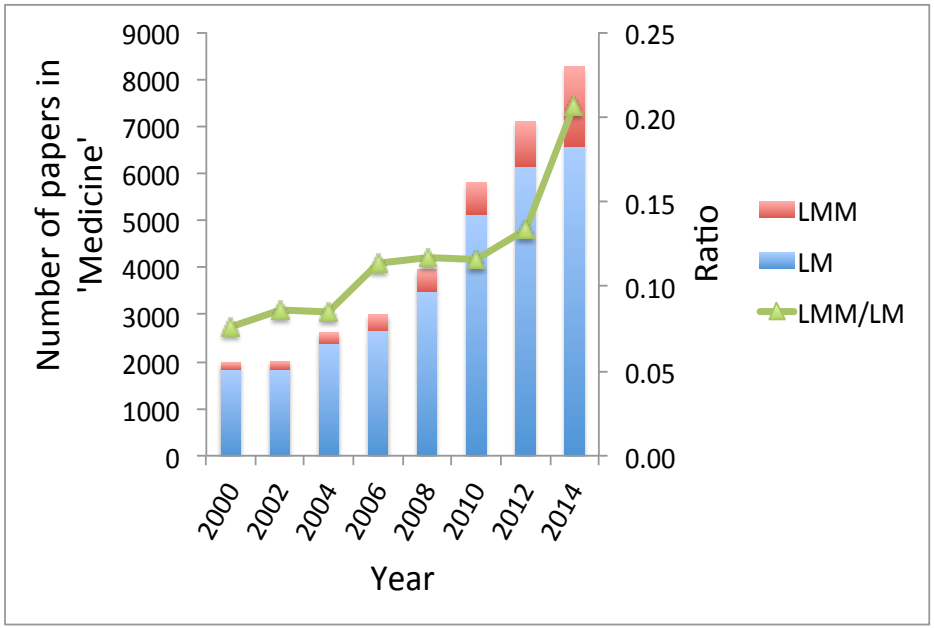
$$\boldsymbol{y} = \mu \mathbf{1}_N + \boldsymbol{Zu} + \boldsymbol{e}$$

$$\boldsymbol{V} = \boldsymbol{Z}(\boldsymbol{A}\sigma_u^2)\boldsymbol{Z}' + \boldsymbol{I}\sigma_e^2$$

- This powerful and flexible tool is now becoming widely used to solve problems in complex traits
  - GWAS, dissecting genetic architecture, system biology, genomic prediction, G x E
- Information (mostly condensed in A and the information matrix) is now substantially different
  - Information based on genome-wide SNPs

Linear mixed model



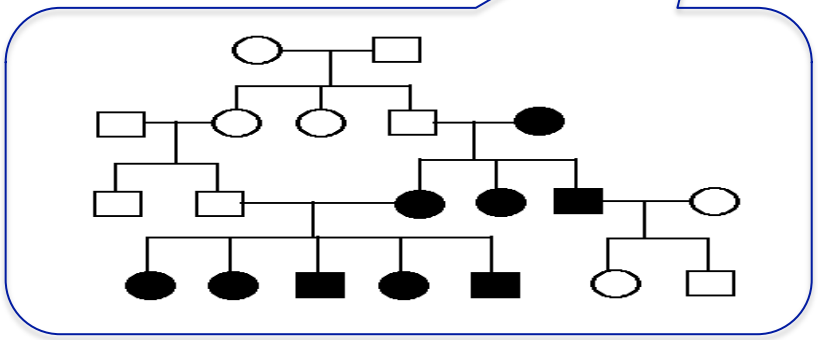


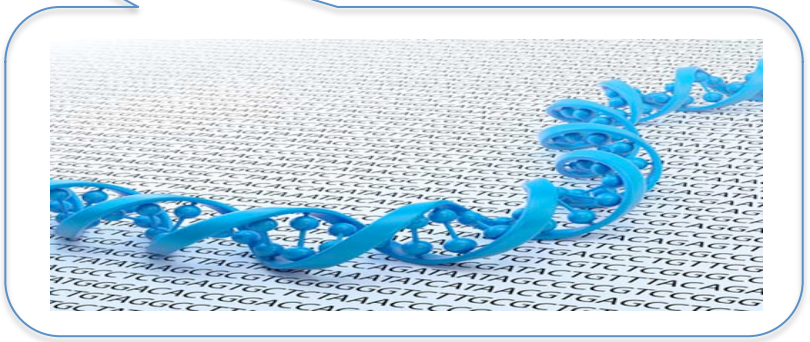
- Rapidly increased in the field (Scopus)

Linear mixed model

$$\boldsymbol{y} = \mu \mathbf{1}_N + \boldsymbol{Zu} + \boldsymbol{e}$$

$$\boldsymbol{V} = \boldsymbol{Z}(\boldsymbol{A}\sigma_u^2)\boldsymbol{Z}' + \boldsymbol{I}\sigma_e^2$$





Traditionally, based on pedigree information

Now, can be based on genomic information

Genomic Relationship Matrix

- GRM can be estimated from *L* SNPs as,

$$\hat{A}_{ij} = \frac{1}{L}\sum_{l=1}^{L}(x_{l[i]} - 2p_l)\cdot(x_{l[j]} - 2p_l)\cdot \mathrm{var}(x_l)^s$$

Van Raden (2008); Yang et al. (2010)

  - where $x_{l[i]}$ is the genotype
  - p is allele frequency
  - var(x_l) is variance of x_l
  - S is the scale parameter (Speed et al. (2012)).

Genomic Relationship Matrix

- Paradigm shifting
  - ✓ Link unrelated individuals through genome-wide SNP similarity
  - ✓ Don't need pedigree, twin design any more
  - ✓ Heritability based on genome-wide SNPs ("SNP-heritability")

Genomic Relationship Matrix

SNP information can also give sample correlation across independent groups

Genomic Relationship Matrix

IQ

Schizophrenia

Possible to estimate genetic correlation between any pair of groups

Genomic Relationship Matrix

Hanwoo

Angus

Possible to estimate genetic correlation between any pair of groups

G x E study

Born in winter

Born in summer

Flexible – don’t need repeated measures
Powerful – sample size can be easily increased

GENOMIC PARTITIONING

Univariate linear mixed model Estimating genetic architecture

ANALYSIS

nature genetics

Common SNPs explain a large proportion of the heritability for human height

Jian Yang[1], Beben Benyamin[1], Brian P McEvoy[1], Scott Gordon[1], Anjali K Henders[1], Dale R Nyholt[1], Pamela A Madden[2], Andrew C Heath[2], Nicholas G Martin[1], Grant W Montgomery[1], Michael E Goddard[3] & Peter M Visscher[1]

- SNP-heritability for human height

Univariate linear mixed model Estimating genetic architecture

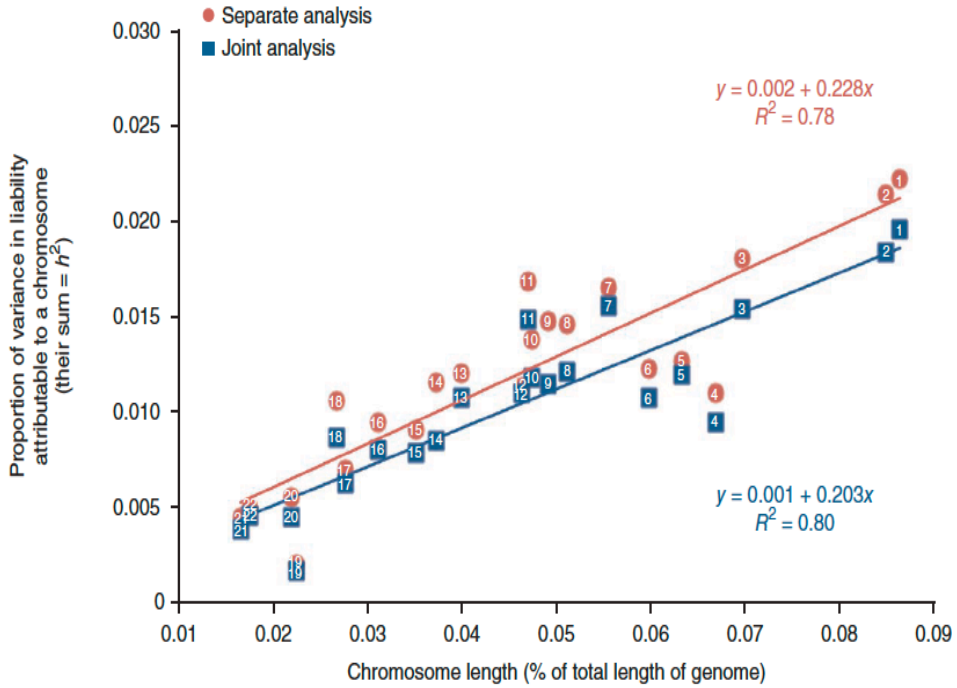
ARTICLE

Estimating Missing Heritability for Disease from Genome-wide Association Studies

Sang Hong Lee,[1] Naomi R. Wray,[1] Michael E. Goddard,[2,3] and Peter M. Visscher[1,*]

- SNP-heritability for human diseases
  - Generalised approach implementing liability threshold model

Genomic partitioning across the genome

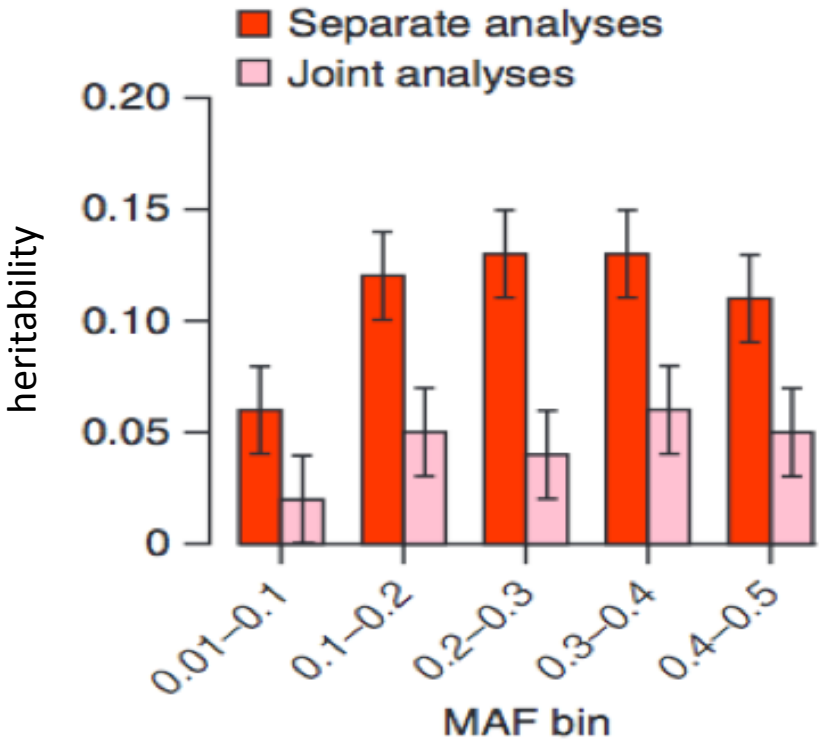


Nat Genet **44**: 247 (2012)

Highly correlated with the length of chromosome

Schizophrenia is polygenic trait (no major genes)

Genomic partitioning across rare/ common variants

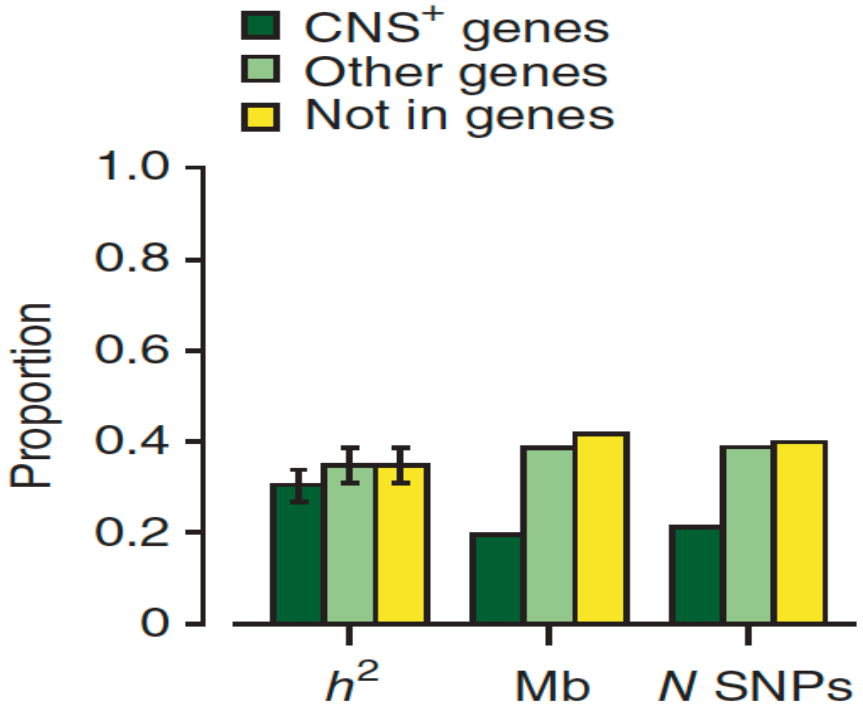


Nat Genet **44**: 247 (2012)

MAF: Minor allele frequency

Schizophrenia caused by common variants

Genomic partitioning across functional category



Nat Genet **44**: 247 (2012)

Bivariate linear mixed model

$$\mathbf{y}_1 = \mathbf{X}_1\mathbf{b}_1 + \mathbf{Z}_1\mathbf{g}_1 + \mathbf{e}_1 \quad \text{for trait 1}$$

$$\mathbf{y}_2 = \mathbf{X}_2\mathbf{b}_2 + \mathbf{Z}_2\mathbf{g}_2 + \mathbf{e}_2 \quad \text{for trait 2}$$

$$\mathbf{V} = \begin{bmatrix} \mathbf{Z}_1\mathbf{A}\mathbf{Z}_1'\sigma^2_{g_1} + \mathbf{I}\sigma^2_{e_1} & \mathbf{Z}_2\mathbf{A}\mathbf{Z}_1'\sigma^2_c \\ \mathbf{Z}_1\mathbf{A}\mathbf{Z}_2'\sigma^2_c & \mathbf{Z}_2\mathbf{A}\mathbf{Z}_2'\sigma^2_{g_2} + \mathbf{I}\sigma^2_{e_2} \end{bmatrix}$$

The y_1 = 0,1 for disorder 1
The y_2 = 0,1 for disorder 2

Y1 and y2 are on different sets of people
The covariance matrix for across all people

BIOINFORMATICS **APPLICATIONS NOTE** *Vol. 28 no. 19 2012, pages 2540–2542*
doi:10.1093/bioinformatics/bts474

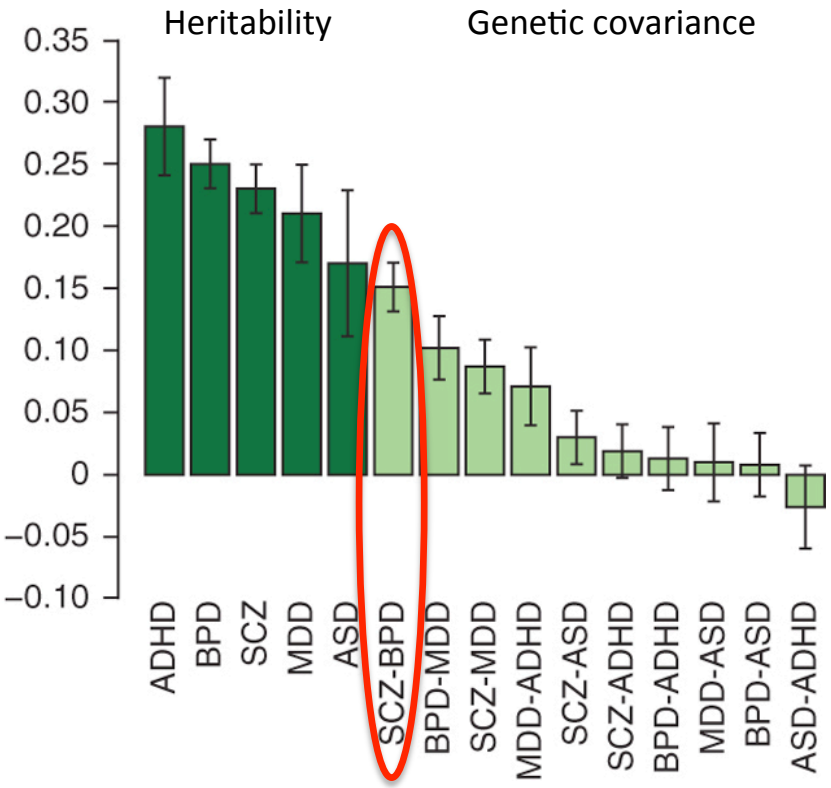
Genetics and population analysis Advance Access publication July 26, 2012

Estimation of pleiotropy between complex diseases using single-nucleotide polymorphism-derived genomic relationships and restricted maximum likelihood

S.H. Lee[1,*], J. Yang[2], M.E. Goddard[3], P.M. Visscher[1,2] and N.R. Wray[1]

Shared genetic architecture (5 psychiatric disorders)

Using bivariate linear mixed model with GRM



Nat Genet **45:** 984 (2013)

SCZ: schizophrenia

BPD: bipolar disorder

MDD: major depressive disorder

ADHD: attention deficit hyperactivity disorder

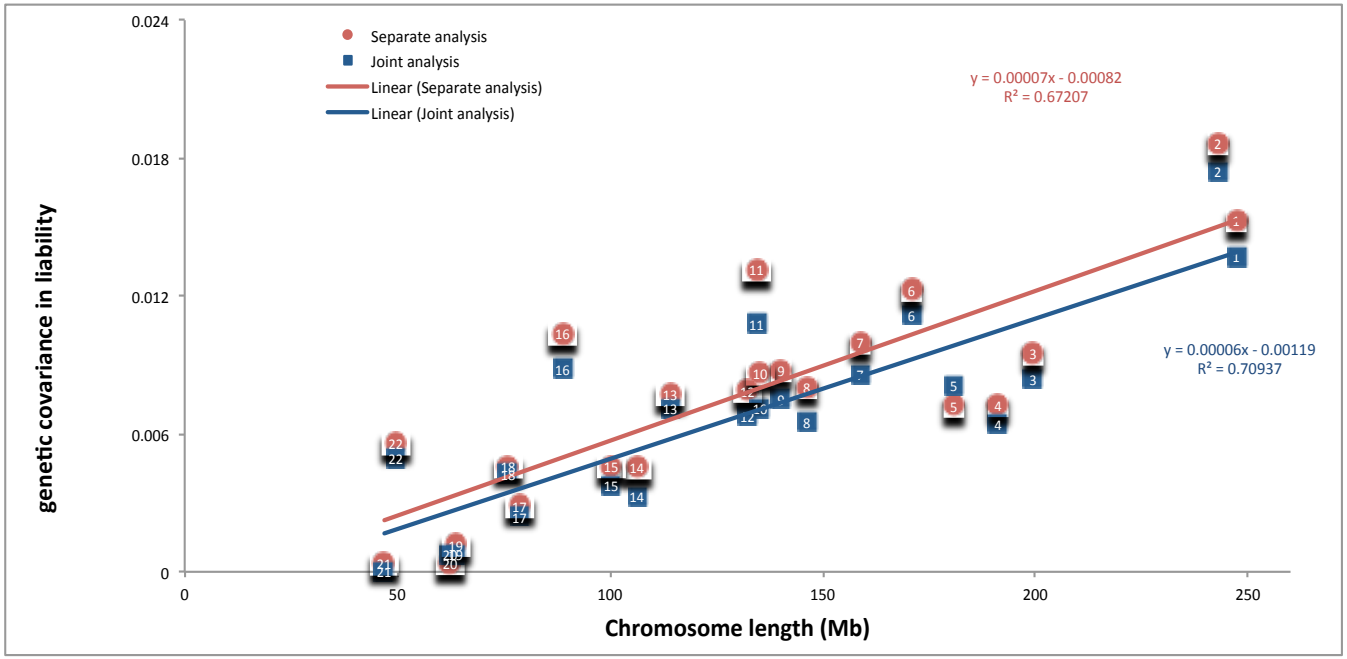
AUT: autism spectrum disorder

SCZ/BIP: high

SCZ/MDD: moderate

SCZ/ASD: low

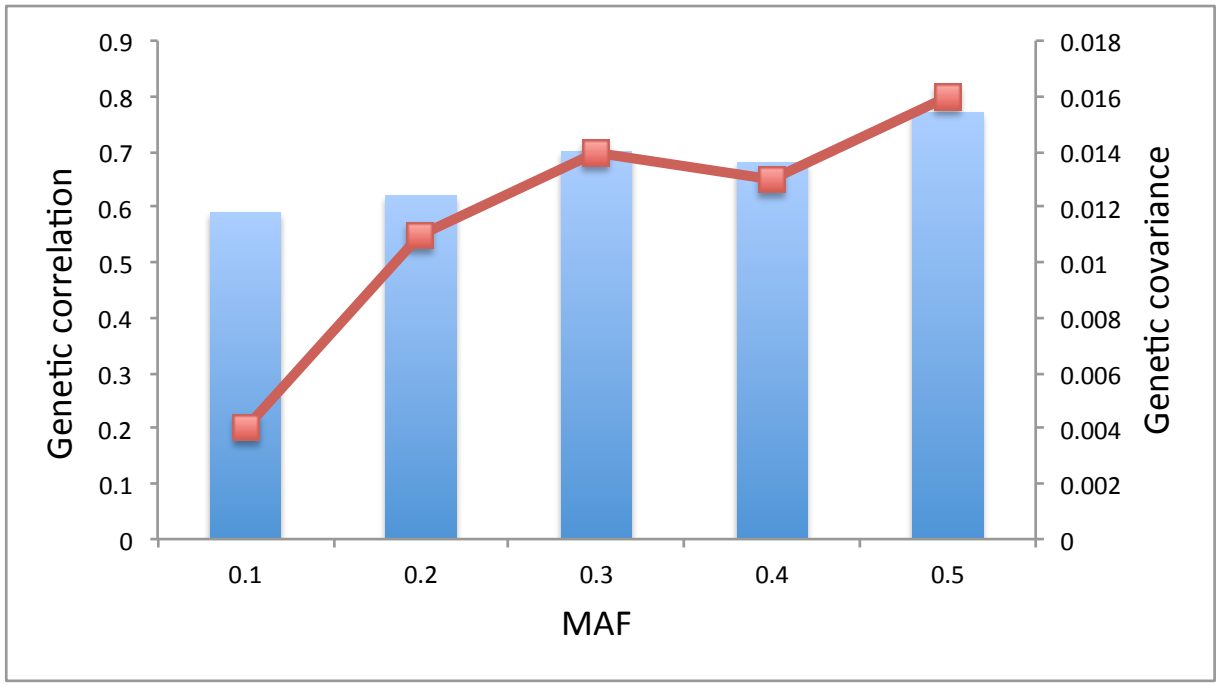
Genetic correlation for SCZ/BIP



Nat Genet **45:** 984 (2013)

- Pleiotropic variants are polygenic

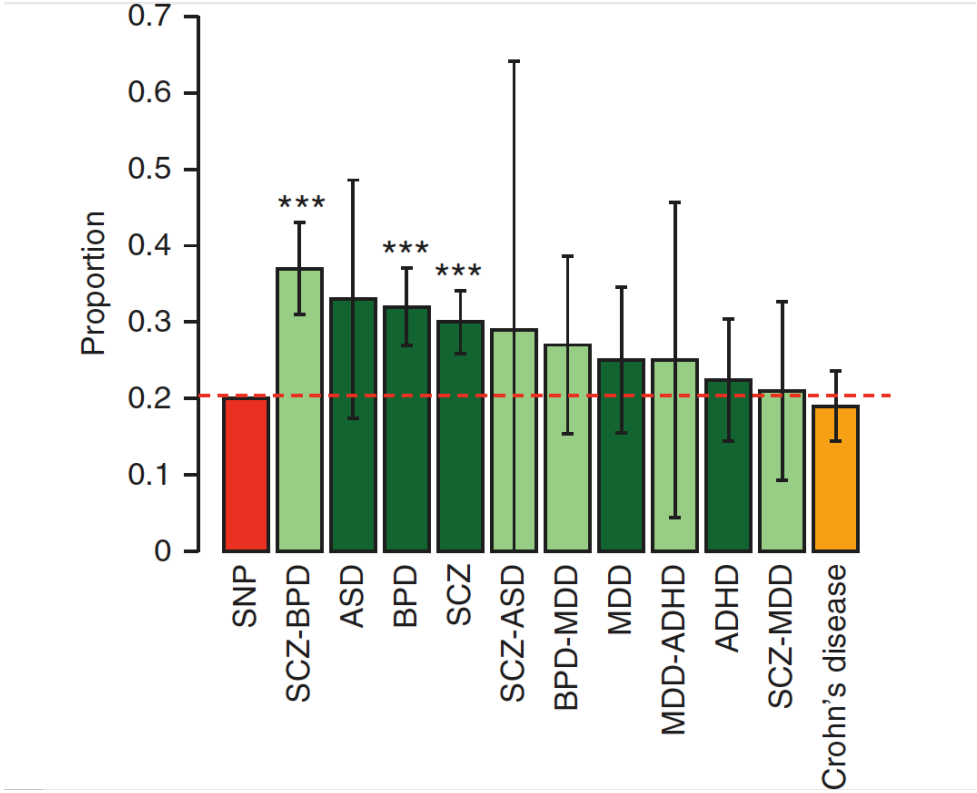
Genetic correlation for SCZ/BIP



Nat Genet **45:** 984 (2013)

- Pleiotropic variants are common (not rare)

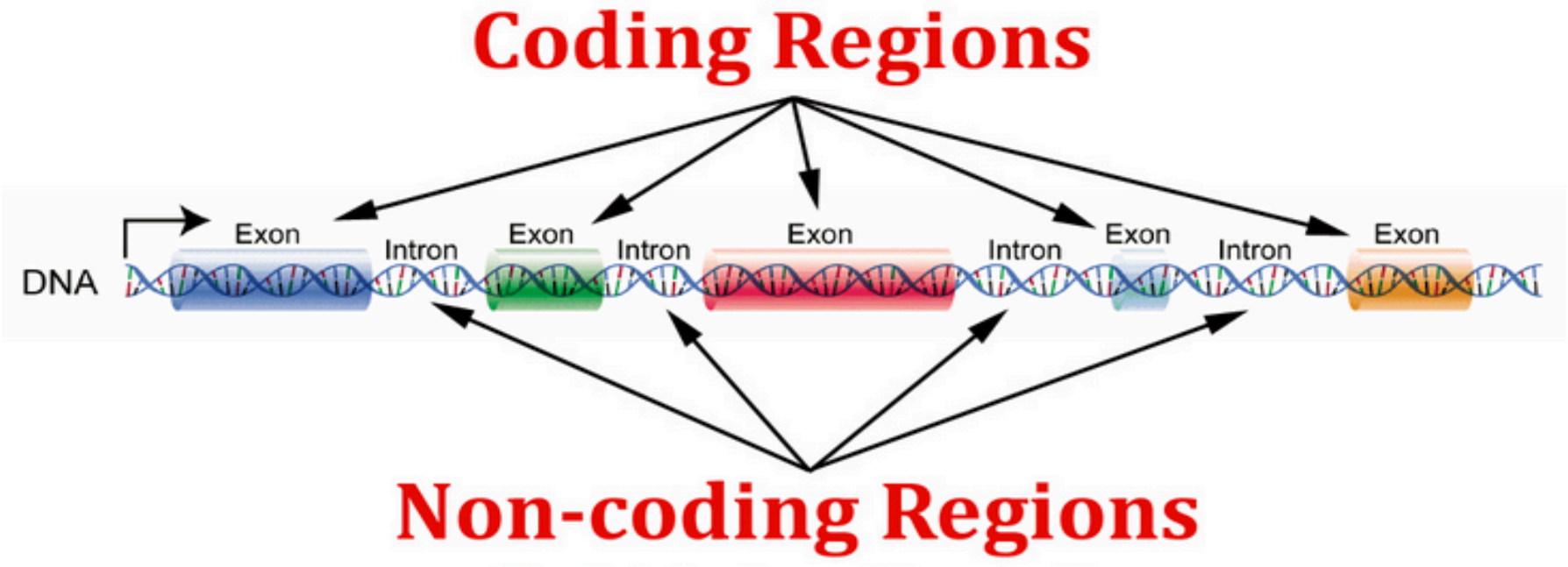
Genetic correlation for SCZ/BIP



Nat Genet **45:** 984 (2013)

Significant disproportion of (co)variance explained by CNS genes

Partitioning G x E



- Magnitude of G x E can be different for coding/non-coding regions

G X E ANALYSES

G x season study

- Swedish SCZ GWAS: 4596 cases/ 5884 controls
  - Winter season (Jan-Apr): 1511 cases/ 2036 controls
  - Non-winter (May-Dec): 2962 cases/ 3772 controls
- SNPs
  - Identify SNPs with association $p < 0.05$ from PGC2
  - Clump pairwise R2 <0.25, 250 kb window
  - MAF > 0.05
  - 47,318 SNPs in total
  - 2,820 SNPs in the C&R (coding & regulatory) set
  - 44,498 SNPs in the All except C&R set

Heterogeneity in coding & regulatory

	h^2 (SE) accounted for by SNPs attributed to:	
	C&R	All SNPs ex C&R
# SNP	2820	44498
SNP-heritability for winter	0.055 (0.010)	0.474 (0.030)
SNP-heritability for non-winter	0.030 (0.006)	0.476 (0.015)
Winter/non-winter SNP correlation	0.564 (0.140)	0.949 (0.048)
p-value	0.00090	0.1459

International Journal of Epidemiology (2015)

- r_g between winter/non-winter significantly less than 1 in C&R,i.e. G x E (Falconer and Mackay 1996)
- No G x E signal in the other region ex C&R

G x maternal environments

- UK Biobank sample
  - ✓40 – 69 years old (recruited 2006 – 2010)
- Genotyped for 150,000 community sample
  - ✓~ 1M common SNPs after stringent QC
- Four later performances phenotypes
  - ✓Fluid intelligence, Memory, react time and educational attainment

G x maternal environments

	Fluid intelligence	Memory	Reaction time	Educational attainment
B&NS	13204	39687	39531	39422
B&S	5311	16093	16021	15990
NB&NS	4362	14015	13937	13919
NB&S	2568	8488	8430	8419
Sum	25445	78283	77919	77750

- B&NS: breastfed and not exposed to maternal smoking around birth,
- B&S: breastfed and exposed to maternal smoking around birth,
- NB&NS: not breastfed and not exposed to maternal smoking around birth,
- NB&S: not breastfed and exposed to maternal smoking around birth

G x maternal environments

	Fluid intelligence	Memory	Reaction time	Educational attainment
Fluid intelligence (N=25445)	0.215 (0.014)	-0.346 (0.046)	-0.186 (0.045)	0.672 (0.028)
Memory (N=78283)	-0.125 (0.006)	0.066 (0.004)	0.126 (0.045)	-0.315 (0.032)
Reaction time (N=77919)	-0.118 (0.006)	0.064 (0.003)	0.080 (0.005)	-0.088 (0.031)
Educational attainment (N=77750)	0.395 (0.005)	-0.115 (0.003)	-0.083 (0.003)	0.177 (0.005)

- Heritability and phenotypic correlation \genetic correlation

G x maternal environments

	Estimate	SE	P-value	
h^2 for B&NS	0.219	0.025	6.5E-19[a]	***
h^2 for B&S	0.260	0.059	9.9E-06[a]	***
h^2 for NB&NS	0.366	0.073	5.9E-07[a]	***
h^2 for NB&S	0.139	0.117	2.3E-01[a]	
r_G (B&S, B&NS)	0.931	0.149	6.4E-01[b]	
r_G (NB&NS, B&NS)	0.597	0.123	1.0E-03[b]	**
r_G (NB&NS, B&S)	0.345	0.159	3.9E-05[b]	***
r_G (NB&S, B&NS)	1.213	0.546	7.0E-01[b]	
r_G (NB&S, B&S)	1.202	0.577	7.3E-01[b]	
r_G (NB&S, NB&NS)	1.060	0.526	9.1E-01[b]	

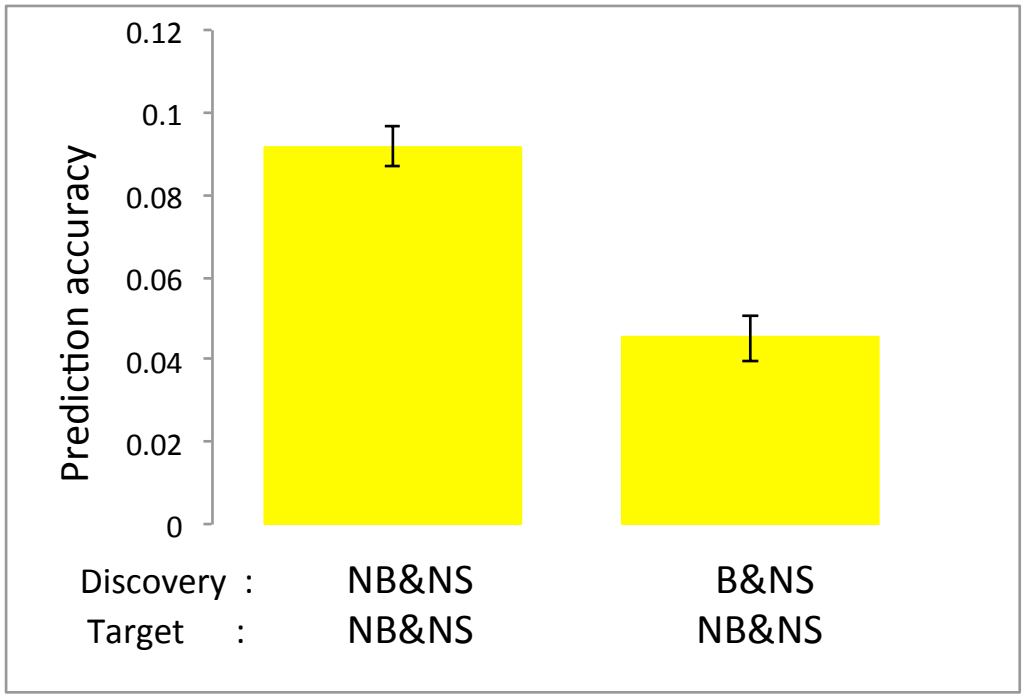
- SNP-heritability and genetic correlation between different E groups for fluid intelligence

G x maternal environments

	Estimate	SE	P-value	
h^2 for B&NS	0.184	0.009	4.0E-87[a]	***
h^2 for B&S	0.169	0.020	4.2E-17[a]	***
h^2 for NB&NS	0.212	0.023	3.9E-20[a]	***
h^2 for NB&S	0.163	0.037	8.4E-06[a]	***
r_G (B&S, B&NS)	0.912	0.073	2.3E-01[b]	
r_G (NB&NS, B&NS)	0.748	0.064	8.3E-05[b]	***
r_G (NB&NS, B&S)	0.866	0.099	1.8E-01[b]	
r_G (NB&S, B&NS)	1.013	0.129	9.2E-01[b]	
r_G (NB&S, B&S)	1.090	0.167	5.9E-01[b]	
r_G (NB&S, NB&NS)	0.929	0.153	6.4E-01[b]	

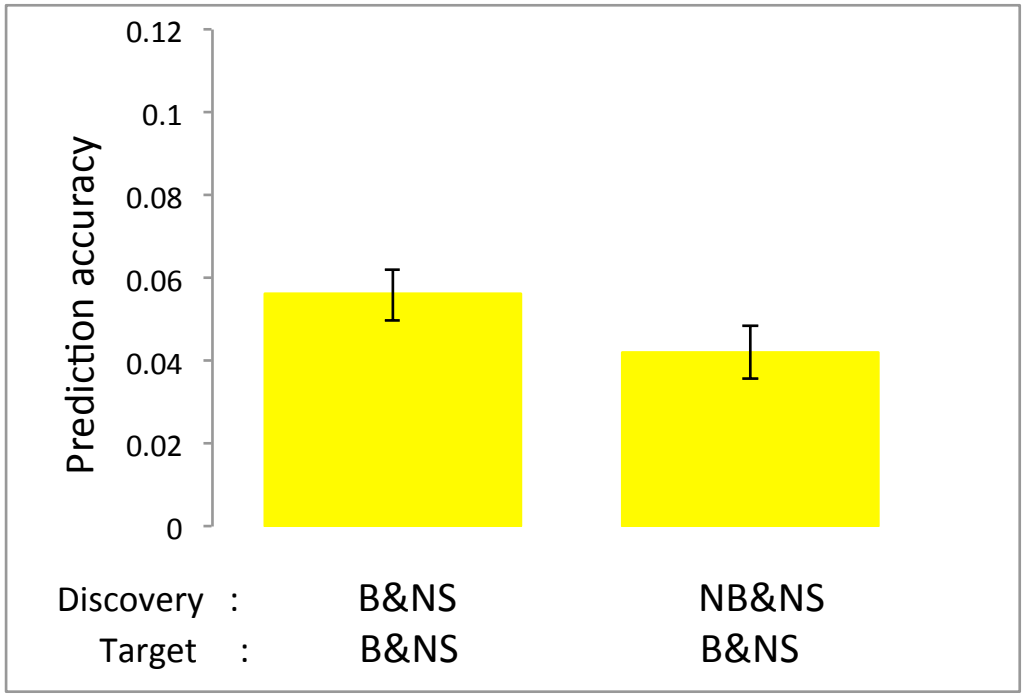
- SNP-heritability and genetic correlation between different E groups for educational attainment

G x E & genomic prediction



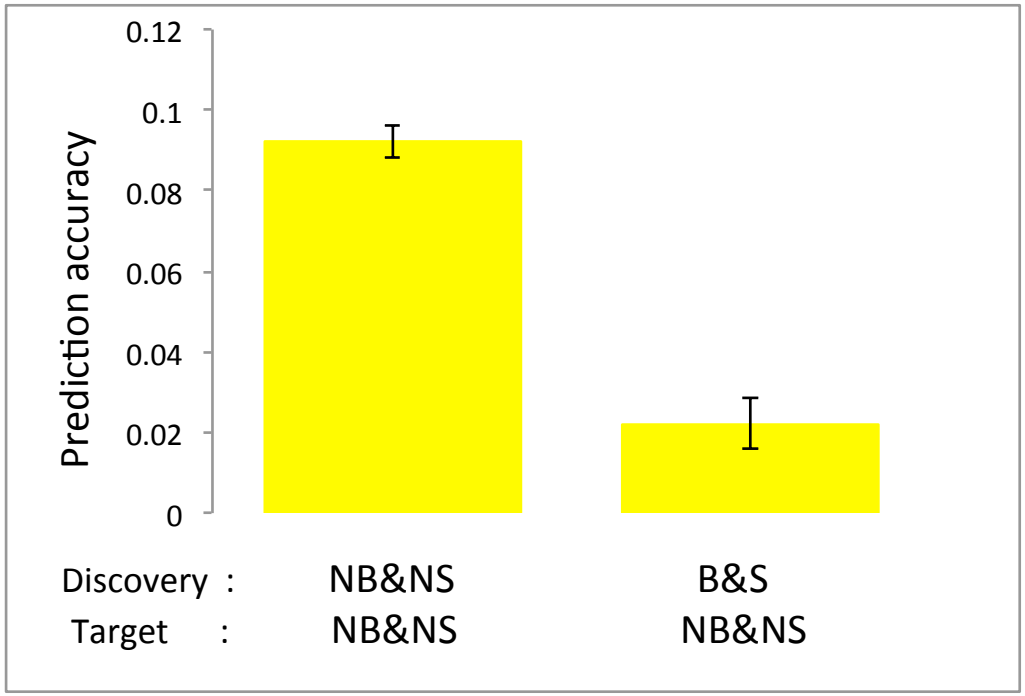
- Fluid intelligence

G x E & genomic prediction



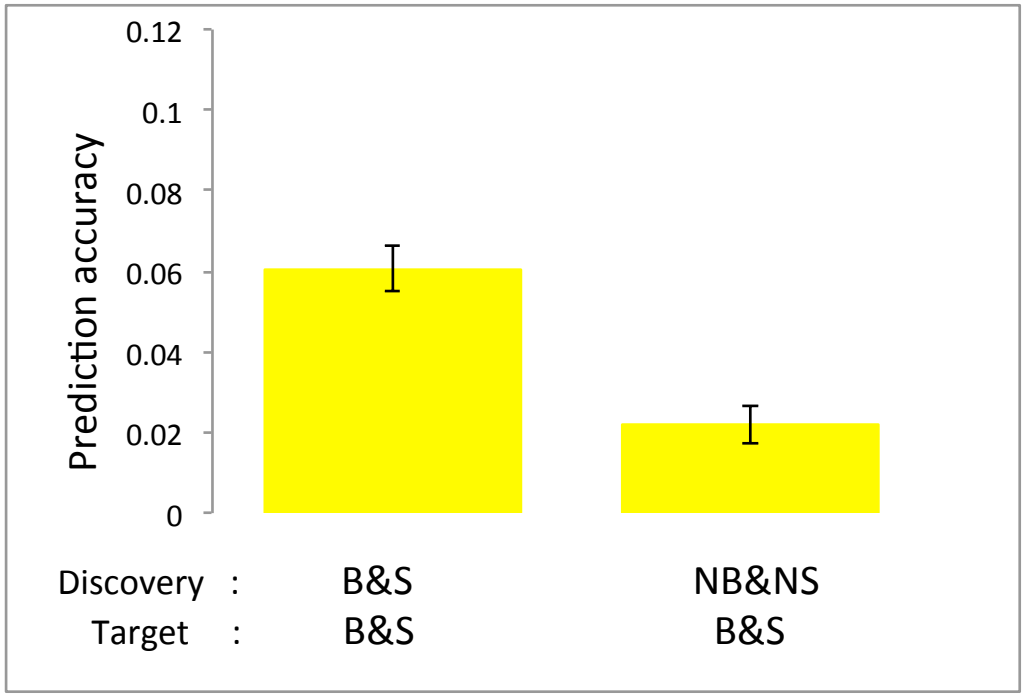
- Fluid intelligence

G x E & genomic prediction



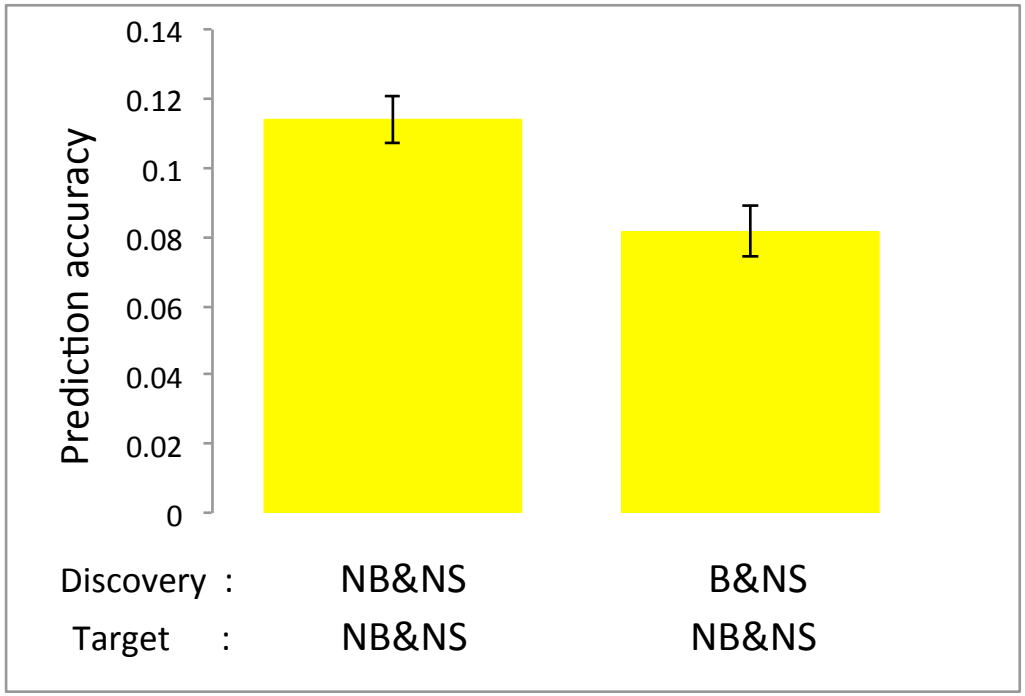
- Fluid intelligence

G x E & genomic prediction



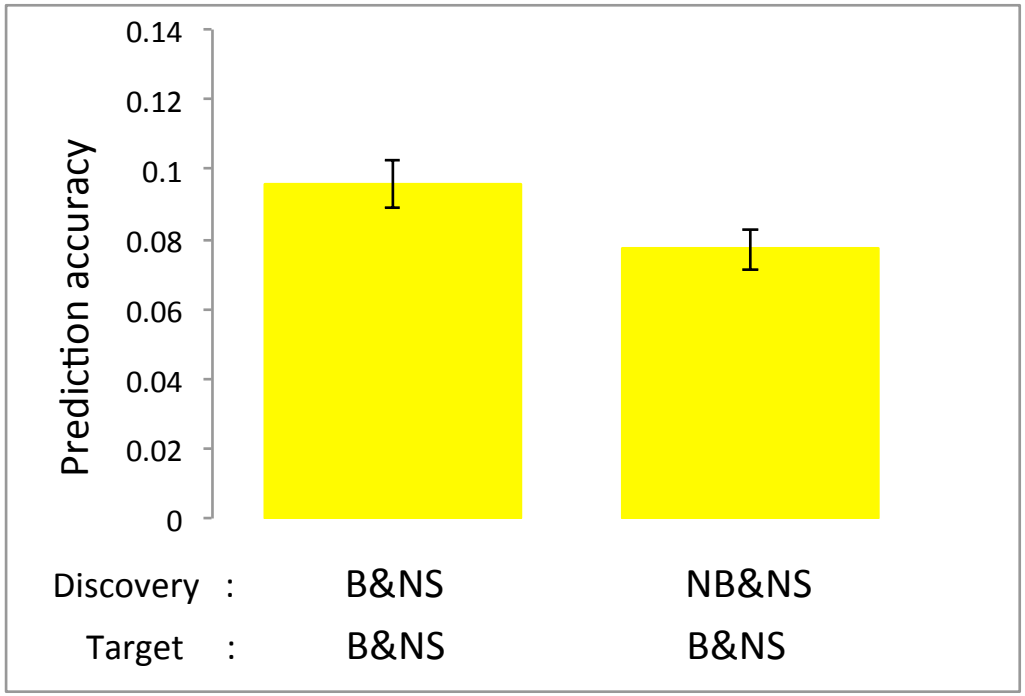
- Fluid intelligence

G x E & genomic prediction



- Educational attainment

G x E & genomic prediction



- Educational attainment

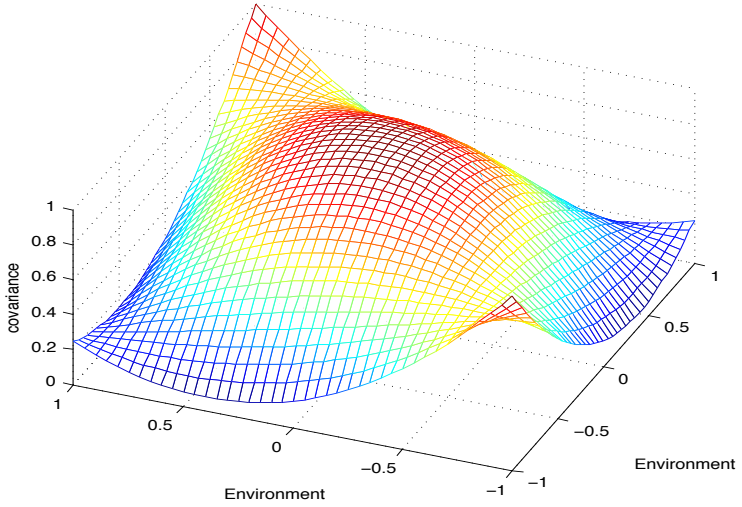
More general G x E model
'Random regression model'

$$y_t = Xb + u_t + e_t$$

$$u_t = \alpha_0 + \alpha_1 \cdot t + \alpha_2 \cdot t^2$$

$$\text{var}\begin{pmatrix} g_1 & \cdots & g_{1,T} \\ \vdots & \ddots & \vdots \\ g_{8,T} & \cdots & g_T \end{pmatrix} = \Phi \mathbf{K} \Phi'$$

$$\mathbf{K} = \begin{bmatrix} \text{var}(\alpha_0) & \text{cov}(\alpha_0,\alpha_1) & \text{cov}(\alpha_0,\alpha_2) \\ \text{cov}(\alpha_0,\alpha_1) & \text{var}(\alpha_1) & \text{cov}(\alpha_1,\alpha_2) \\ \text{cov}(\alpha_0,\alpha_2) & \text{cov}(\alpha_1,\alpha_2) & \text{var}(\alpha_2) \end{bmatrix}$$

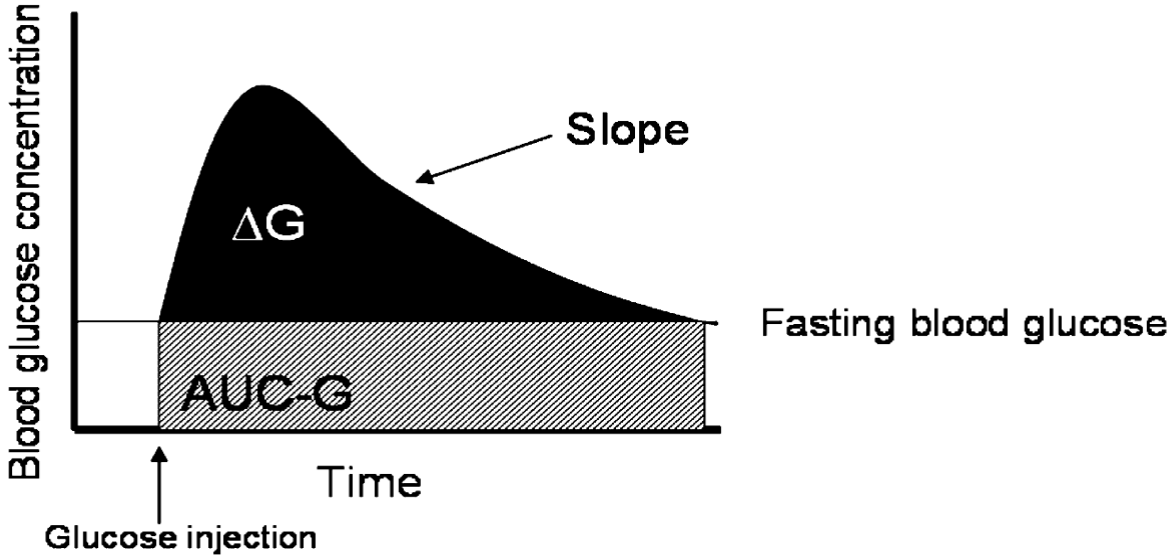


- t – value for environment (age, season)
- T – # observed points (simulated as 4)
- α – random regression coefficients
- ϕ – Legendre polynomials with t
- K – coefficient matrix
- k – order of the matrix

Real data

- Heterogeneous stock mice data
- After a stringent QC, we used 9,258 autosomal SNPs from 1,908 individuals
- phenotypes of four glucose values (taken at 0, 15, 30 and 75 minutes after intraperitoneal glucose injection in a model of type 2 diabetes mellitus
- Estimate G x E for glucose levels if there is interaction

Glucose level trajectory



Solberg et al. (2006) Mammalian Genome **17:** 129 - 46

- Glucose level measured at 0, 15, 30 and 75 minutes from the injection

Real data

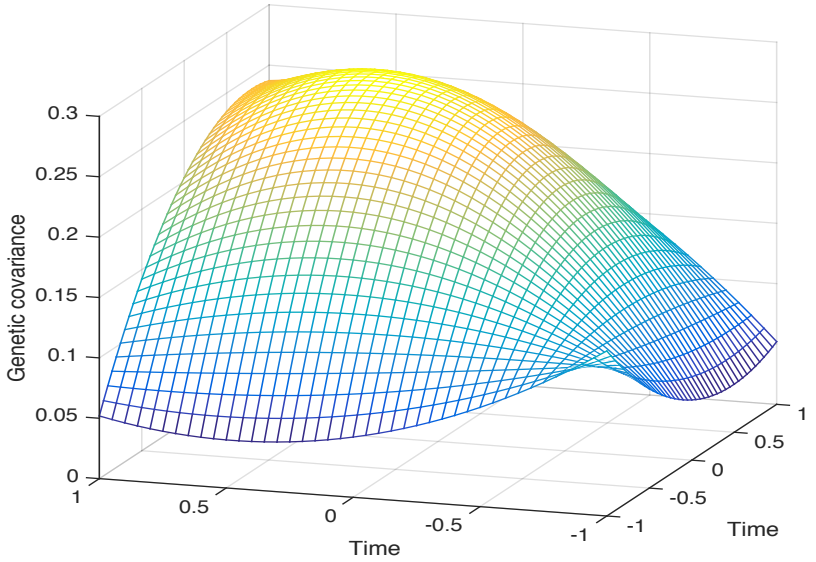
		# Order for residual variance		
		1	2	3
# Order for genetic variance	1	-7363.56	-7708.08	-8471.70
	2	-7700.28	-7854.00	-8606.18
	3	-7929.46	-8110.98	-8636.36
	4	-7939.51	-8123.02	-8623.00

Bioinformatics **32:** 1420 - 2 (2016)

- Significant genetic heterogeneity across the trajectory, i.e. evidence of G x E

Real data

$$\text{var}\begin{pmatrix} g_1 & \cdots & g_{1,T} \\ \vdots & \ddots & \vdots \\ g_{8,T} & \cdots & g_T \end{pmatrix} = \Phi \mathbf{K} \Phi'$$



- Variance covariance structure can be obtained from estimated variance of random regression coefficients (**K**) and Legendre polynomial function (ϕ).

ALGORITHM

Algorithm for REML

- Average Information algorithm
  - Average of Hessian and Fisher's matrix

$$\Theta^{(k+1)} = \Theta^{(k)} - (\mathbf{H}^{(k)})^{-1} \frac{\partial L}{\partial \Theta}\Big|\Theta^{(k)} \qquad \Theta^{(k+1)} = \Theta^{(k)} + (F^{(k)})^{-1} \frac{\partial L}{\partial \Theta}\Big|\Theta^{(k)} \qquad \Theta^{(k+1)} = \Theta^{(k)} + (\mathbf{AI}^{(k)})^{-1} \frac{\partial L}{\partial \Theta}\Big|\Theta^{(k)}$$

$$\mathbf{H} = \frac{\partial^2 L}{\partial \sigma_i^2 \partial \sigma_j^2} = \frac{1}{2}\left[tr\left(\frac{\partial \mathbf{V}}{\partial \sigma_i^2}\mathbf{P}\frac{\partial \mathbf{V}}{\partial \sigma_j^2}\mathbf{P} \right) - \mathbf{y}'\frac{\partial \mathbf{V}}{\partial \sigma_i^2}\mathbf{P}\frac{\partial \mathbf{V}}{\partial \sigma_j^2}\mathbf{PPy} \right]$$

$$\mathbf{F} = E\left(\frac{\partial^2 L}{\partial \sigma_i^2 \partial \sigma_j^2} \right) = \frac{1}{2}\left[tr\left(\frac{\partial \mathbf{V}}{\partial \sigma_i^2}\mathbf{P}\frac{\partial \mathbf{V}}{\partial \sigma_j^2}\mathbf{P} \right) \right]$$

$$\mathbf{AI} = \frac{1}{2}\left[\mathbf{y}'\frac{\partial \mathbf{V}}{\partial \sigma_i^2}\mathbf{P}\frac{\partial \mathbf{V}}{\partial \sigma_j^2}\mathbf{PPy} \right]$$

Algorithm for REML

- Average Information algorithm
  - Based on Mixed model equation (Gilmour et al. 1995)

$$\begin{bmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{X}'\mathbf{R}^{-1}\mathbf{Z} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z}+\mathbf{G}^{-1} \end{bmatrix} \begin{bmatrix} \hat{\mathbf{b}} \\ \hat{\mathbf{u}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{y} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{y} \end{bmatrix}$$

  - based on V matrix (Lee and van der Werf 2006)

$$\mathbf{V} = \mathbf{Z}\mathbf{G}\mathbf{Z}' + \mathbf{R}$$

$$\hat{\mathbf{b}} = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{y})$$

$$\hat{\mathbf{u}} = \mathbf{G}\mathbf{Z}'\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\mathbf{b})$$

Algorithm for REML

- **Direct AIREML (based on V matrix)**
  - ✓ Efficient for dense GRM
  - ✓ Efficient for multiple random effects (e,g, chromosome, MAF)
  - ✓ Compared to ASReml, computational efficiency > 100 times
  - ✓ Implemented in GCTA, multiBLUP, MTG2

$$\begin{bmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{X}'\mathbf{R}^{-1}\mathbf{Z}_1 & \cdots & \mathbf{X}'\mathbf{R}^{-1}\mathbf{Z}_N \\ \mathbf{Z}_1'\mathbf{R}^{-1}\mathbf{X} & \mathbf{Z}_1'\mathbf{R}^{-1}\mathbf{Z}_1 + \mathbf{G}_1^{-1} & \cdots & \mathbf{Z}_1'\mathbf{R}^{-1}\mathbf{Z}_N \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{Z}_N'\mathbf{R}^{-1}\mathbf{X} & \mathbf{Z}_N'\mathbf{R}^{-1}\mathbf{Z}_1 & \cdots & \mathbf{Z}_N'\mathbf{R}^{-1}\mathbf{Z}_N + \mathbf{G}_N^{-1} \end{bmatrix} \begin{bmatrix} \hat{\mathbf{b}} \\ \mathbf{u}_1 \\ \vdots \\ \mathbf{u}_N \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{y} \\ \mathbf{Z}_1\mathbf{R}^{-1}\mathbf{y} \\ \vdots \\ \mathbf{Z}_N\mathbf{R}^{-1}\mathbf{y} \end{bmatrix}$$

$$\mathbf{V} = \mathbf{Z}_1\mathbf{G}_1\mathbf{Z}_1' + \ldots + \mathbf{Z}_N\mathbf{G}_N\mathbf{Z}_N + \mathbf{R}$$

Algorithm for REML

- Direct AI algorithm (Lee and van der Werf 2006)

$$\mathbf{V} = \mathrm{var}(\mathbf{y}_i) = \begin{bmatrix} \mathbf{Z}_1\mathbf{A}\sigma^2_{g_1}\mathbf{Z}'_1 + \mathbf{Z}_1\mathbf{I}\sigma^2_{e_1}\mathbf{Z}'_1 & \cdots & \mathbf{Z}_1\mathbf{A}\sigma_{g_{1,t}}\mathbf{Z}'_t + \mathbf{Z}_1\mathbf{I}\sigma_{e_{1,t}}\mathbf{Z}'_t \\ \vdots & \ddots & \vdots \\ \mathbf{Z}_1\mathbf{A}\sigma_{g_{1,t}}\mathbf{Z}'_t + \mathbf{Z}_1\mathbf{I}\sigma_{e_{1,t}}\mathbf{Z}'_t & \cdots & \mathbf{Z}_t\mathbf{A}\sigma^2_{g_t}\mathbf{Z}'_t + \mathbf{Z}_t\mathbf{I}\sigma^2_{e_t}\mathbf{Z}'_t \end{bmatrix}$$

The relationship matrix can be decomposed as **A=UDU'** where **I=UU'** (Thompson and Shaw, 1990)

$$\mathbf{U}'\mathbf{y}_i = \mathbf{U}'\mathbf{X}_i\mathbf{b}_i + \mathbf{U}'\mathbf{Z}_i\mathbf{g}_i + \mathbf{U}'\mathbf{e}_i$$

$$\mathbf{V} = \mathrm{var}(\mathbf{U}'\mathbf{y}_i) = \begin{bmatrix} \mathbf{D}\sigma^2_{g_1} + \mathbf{I}\sigma^2_{e_1} & \cdots & \mathbf{D}\sigma_{g_{1,n}} + \mathbf{I}\sigma_{e_{1,n}} \\ \vdots & \ddots & \vdots \\ \mathbf{D}\sigma_{g_{n,1}} + \mathbf{I}\sigma_{e_{n,1}} & \cdots & \mathbf{D}\sigma^2_{g_n} + \mathbf{I}\sigma^2_{e_n} \end{bmatrix}$$

Algorithm for REML

- Real data application
  - Heterogeneous stock mice data (WTC)
  - After a stringent QC of genotypic data, we used 9,258 autosomal SNP from 1,908 individuals
    - Construct GRM based on the SNPs
  - Four glucose values (taken at 0, 15, 30 and 75 minutes after intraperitoneal glucose injection in a model of type 2 diabetes mellitus) and body mass index (BMI).
  - Five-trait linear mixed model and random regression model for the repeated glucose measures.

Results

	MTG2	GEMMA	ASReml	WOMBAT
# traits		Multivariate linear mixed model		
1	1 sec	1 sec	2 min	17 sec
3	1 sec	1 sec	210 min	9 min
5	2 sec	6 sec	950 min	60 min
# order		Random regression model		
1	2 sec	NA[a]	4 min	3 min
2	2 sec	NA	82 min	30 min
3	2 sec	NA	310 min	54 min

- With eigen-decomposition trick
  - Single genetic covariance structure (GRM)
  - Computational efficiency increase > 1000 times

Results

	# Random effects		
	1	2	3
	Multivariate linear mixed model (3 traits model)		
MTG2	6min	8min	9min
ASReml	210min	1300 min	4200min
WOMBAT	9min	62min	130min
	Random regression model		
	(# Order=2 for genetic and residual variance component)		
MTG2	27min	39min	63min
ASReml	82min	470min	> 1000min
WOMBAT	30min	80min	250min

- Without eigen-decomposition trick
  - Possible to fit multiple GRMs
  - MTG2 still performs better than others

Results

	Multivariate linear mixed model	Random regression model
	With a single genetic covariance matrix	
MTG2	$O(n^3 + n^2t + n^2c + r_1nt^3 + r_2nct^6)$	$O(n^3 + n^2p + n^2c + r_1np^3 + r_2ncp^2k^4)$
GEMMA	$O(n^3 + n^2t + n^2c + r_3nc^2t^2 + r_2nc^2t^6)$	NA
ASReml	$O(r_3n^3(t + c)^3 + r_2n^3t^7)$	$O(r_3n^3(p + c)^3 + r_2n^3p^3k^4)$
WOMBAT	$O(r_3n^3(t + c)^3 + r_2n^3t^7)$	$O(r_3n^3(p + c)^3 + r_2n^3p^3k^4)$
	With multiple genetic covariance matrices (m>1)	
MTG2	$O(r_1n^3t^3 + r_2n^3t^7m^4)$	$O(r_1n^3p^3 + r_2n^3p^3k^4\ m^4)$
GEMMA	NA	NA
ASReml	$O(r_3n^3\ m^3\ (t + c)^3 + r_2n^3t^7m^7)$	$O(r_3n^3\ m^3\ (p + c)^3 + r_2n^3p^3k^4m^7)$
WOMBAT	$O(r_3n^3\ m^3\ (t + c)^3 + r_2n^3t^7m^7)$	$O(r_3n^3\ m^3\ (p + c)^3 + r_2n^3p^3k^4m^7)$

Approximated computational complexity of the methods

Results

Replicates	MTG2					ASReml[c]		Converged
	Likelihood	# Iteration				Likelihood	# Iteration (r_2^e)	
		Poor starting value[a]		Good starting value[b]				
		r_1^d	r_2^e	r_1^d	r_2^e			
1	-4424.15	2	6	0	3	-4424.15	9	yes
2	-4378.5	11	10	0	4	-4378.5	8	yes
3	-4375.63	3	6	0	4	-4375.63	8	yes
4	-4326.18	3	6	0	3	-4326.18	8	yes
5	-4354.97	3	6	0	3	-4354.97	15	yes
6	-4411.81	2	6	0	4	-4311.81	14	yes
7	-4093.76	2	5	0	3	-4093.76	9	yes
8	-4216.42	2	6	0	4	-4216.42	10	yes
9	-4119.63	2	5	0	3	-4119.63	8	yes
10	-4218.67	2	6	0	4	-4218.67	10	yes

Three-trait linear mixed model for the heterogeneous mice stock

✓ Identical results from MTG2 and ASReml

Results

Replicates	MTG2					ASReml		Converged
	Likelihood	# Iteration				Likelihood	# Iteration (r_2)	
		Poor starting value		Good starting value				
		r_1	r_2	r_1	r_2			
1	-8324.47	0	9	0	3	-8324.47	7	yes
2	-8311.27	0	8	0	3	-8311.27	7	yes
3	-8259.34	0	8	0	3	-8259.34	7	yes
4	-8325.44	0	8	0	3	-8325.44	7	yes
5	-8315.36	0	8	0	3	-8315.36	7	yes
6	-8292.39	0	8	0	3	-8392.39	7	yes
7	-8367.17	0	8	0	3	-8367.17	7	yes
8	-8363.37	0	8	0	3	-8363.37	7	yes
9	-8329.45	0	8	0	3	-8429.45	8	yes
10	-8371.66	0	8	0	3	-8371.66	7	yes

Random regression model with the Legendre polynomial order of 3 for the genetic and residual variance components

✓ Identical results from MTG2 and ASReml

Discussion

- There are two limitations to MTG2 as well as GEMMA
- The eigen-decomposition technique cannot be used with more than one GRM
  - In models with multiple GRMs, GEMMA cannot be used and MTG2 becomes slow although it is still considerably faster than ASReml and WOMBAT
- Secondly, the eigen-decomposition technique requires complete data across multiple traits (no missing value)
  - Phenotypic imputation can be used
  - EM algorithm (to be implemented)
- MTG2 and WOMBAT can facilitate a parallel computation that increases the efficiency further

MTG2 software

https://sites.google.com/site/honglee0707/mtg2

- mtg2 binary for Linux, window and mac
- manual
- Examples
  - Example0/
  - Example1/
  - Example5/

MTG2 software

- Multivariate linear mixed model
  - MTG-REML
  - MTG-BLUP
    - Individual solution <-> SNP solution
- Random regression linear mixed model
  - Legendre polynomial
  - Spline

MTG2 extra options

-cove 1: parameterising residual covariance

-thread k: k paralleled computation

-sv {file name}

-mg {file name} instead of -g {file name}: multiple random effects model

-bv {file name}: BLUP estimation

-inv 1: inverting matrix

-bend 1: bending NPD matrix making it PD

-nit {value}: maximum number of iterations (default is 200)

-conv {value}: convergence criteria for log likelihood (default is 0.001)

-frq 1: estimating allele frequency given plink files

MTG2 extra options

-trf_h2: transform between observed and liability scale for h2

-sbv: converting between individual BLUP and SNP BLUP

-delta: Delta method to get SE of the ratio for target VC

-rtmx: estimating GRM

-Me: effective # chromosome segments given effective size

-pred_acc: accuracy of genomic prediction based on theory
(and power of genomic prediction)

-simcoal: coalescence simulation (SNP genotype)

-simreal: phenotype simulation based on given genotypes

-fix: constrain some parameters during REML estimation