

Extending GCTA: Effects of the Genetic Environment

2014 International Workshop on Statistical
Genetic Methods
for Human Complex Traits
Boulder, CO.

Lindon Eaves, David Evans

Thanks also:

- Beate St. Pourcin – Bristol U.
- Tim York - VCU
- George Davey-Smith – Bristol U
- Mike Neale - VCU
- Peter Visscher and Matthew Robinson – UQLD

See: Eaves LJ, St. Pourcain B, Davey-Smith G, York TP, Evans DM. (2014) Resolving the effects of maternal and offspring genotype on dyadic outcomes in Genome Wide Complex Trait Analysis (“M-GCTA”). *Behavior Genetics* (under revision).

Funding

“We have no money. We will have to think.”
(Rutherford)

University of Bristol UK:

Benjamin Meaker Visiting Professorship

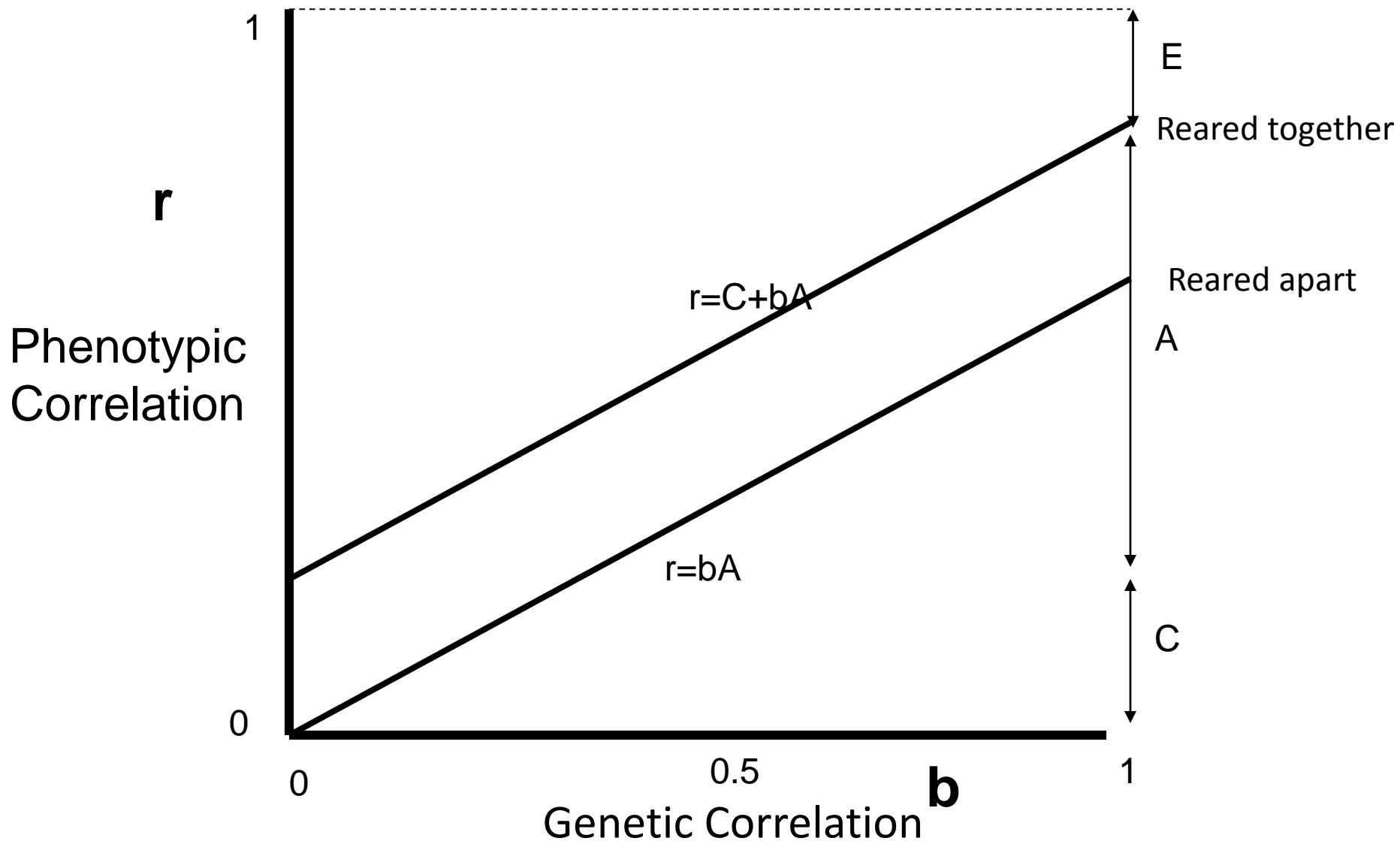
Virginia Retirement System

United States Social Security Administration

“Quantitative Genetics”

Estimates genetic (and environmental) contribution to variation by regressing measures of phenotypic similarity on genetic (and environmental) similarity

Regression of phenotypic correlation on genetic correlation (modified from Fulker, 1979)



Classical Approach (Fisher, Wright, Falconer, Mather etc.)

Estimates genetic correlation from expected genetic resemblance based on genetic relatedness (MZ, DZ twins, siblings, half-siblings, cousins etc.)

“GCTA”

“Genome-Wide Complex Trait Analysis”

Visscher, Yang, Lee, Goddard

Estimates genetic correlation from empirical genetic resemblance of “unrelated” individuals based on identity by state across using genome-wide single nucleotide polymorphisms (“SNPs”)

Basic GCTA Model

$$V = G + E$$

V=Phenotypic variance between individuals

G=(Additive) genetic variance

E=(Unique) environmental variance

$$\Sigma = \mathbf{A}G + \mathbf{I}E$$

Where Σ = expected phenotypic covariance matrix between individuals; \mathbf{A} = empirical genetic relatedness matrix between individuals

GCTA

Plusses

- Doesn't need special samples (e.g. twins)
- Aggregate across genome
- Uses all possible pairs of unrelateds (smallish Ns yield lots of pairs)
- Uses real genomic information to estimate genetic resemblance
- Can partition genetic effects among e.g. chromosomes
- No problem with "special twin environments"

Minusses

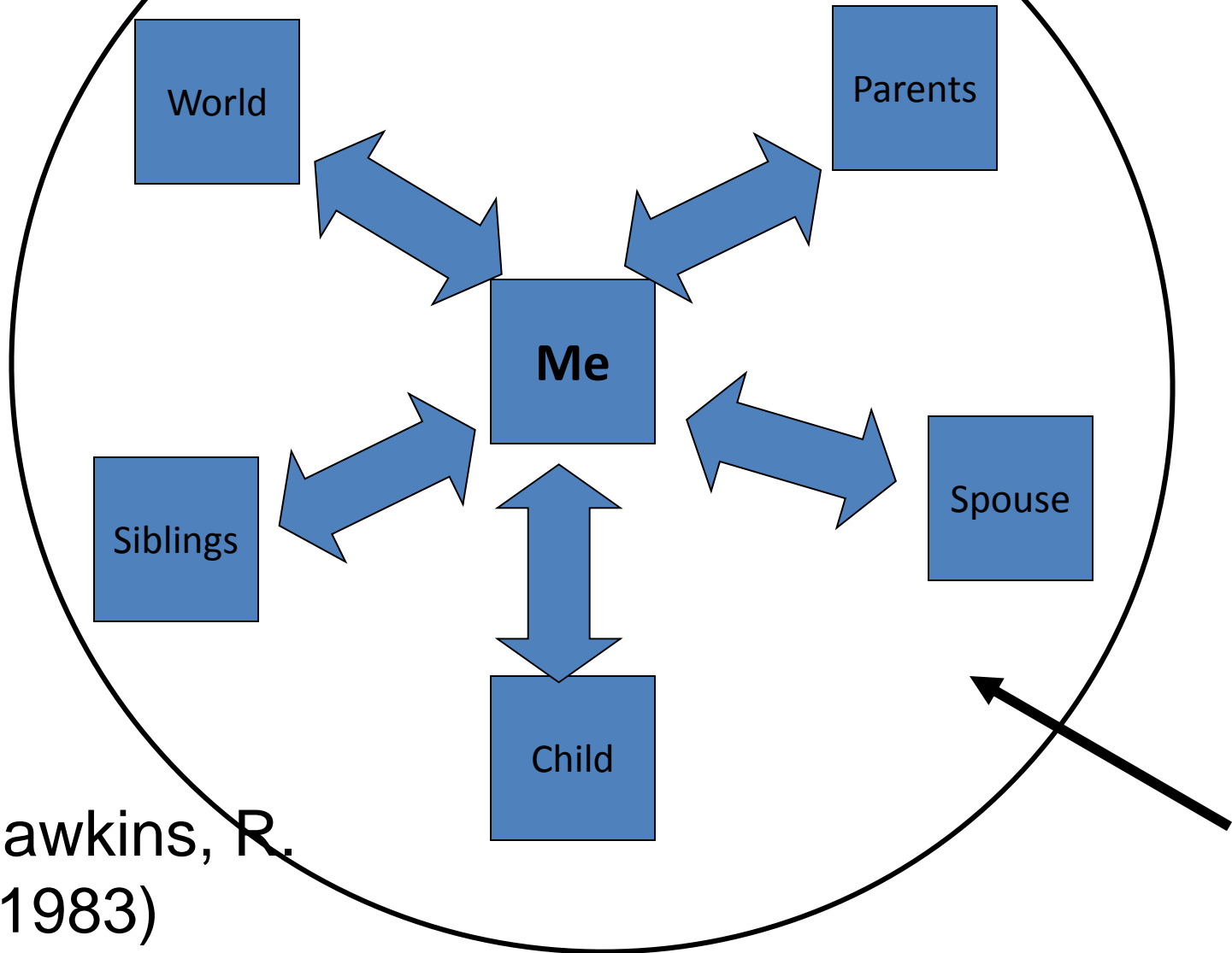
- Ignores lots of stuff at this point (social interaction, non-additive genetic effects)
- Assumes SNPs yield unbiased estimate of genetic correlation (? assortative mating)
- Estimates have large s.e.'s
- Lots of other stuff still being figured out (environmental stratification.....?)

BUT....

Humans are “More than”
individual packets of DNA – have
families, feet, brains and
communities

Behavioral phenotype extends
beyond individual

Extending the Phenotype



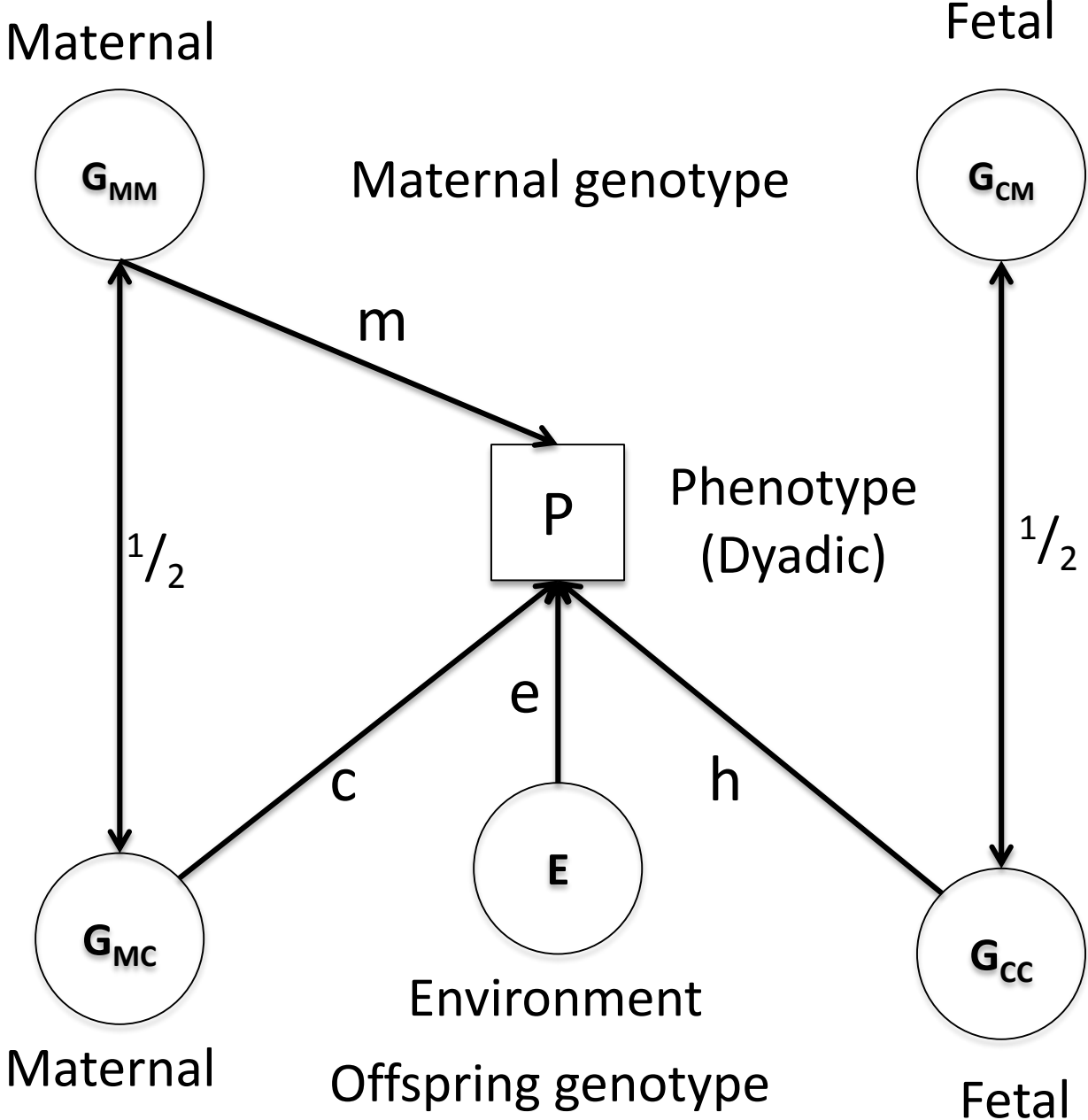
c.f. Dawkins, R.
(1983)

A Place to Start: M-GCTA

Include effects of the Maternal Genotype on Offspring Behavior (“Genetic Maternal Effects”)

- Extensive background in plant and animal genetics (reciprocal crosses, diallels, dam and sire effects)
- Good models and examples from human studies (extended kinships of twins, maternal and paternal half-sibships)
- Extensively measured and genotyped mother-child pairs (e.g. “ALSPAC”)

Basic path model for fetal and maternal effects in mother-child dyads.



Basic model for components of variance

$$V = h^2 + m^2 + mc + e^2 \text{ (path model)}$$

Or:

$$V = G + M + Q + E$$

(Biometrical-genetic model where:

$$M=m^2, G=(c^2 + h^2) \text{ and } Q = mc)$$

V=Phenotypic variance

G=(Additive) genetic variance (“A” in ACE model)

M=Shared environmental variance due to indirect effects of maternal genotype on offspring phenotype (“genetic maternal effects”)

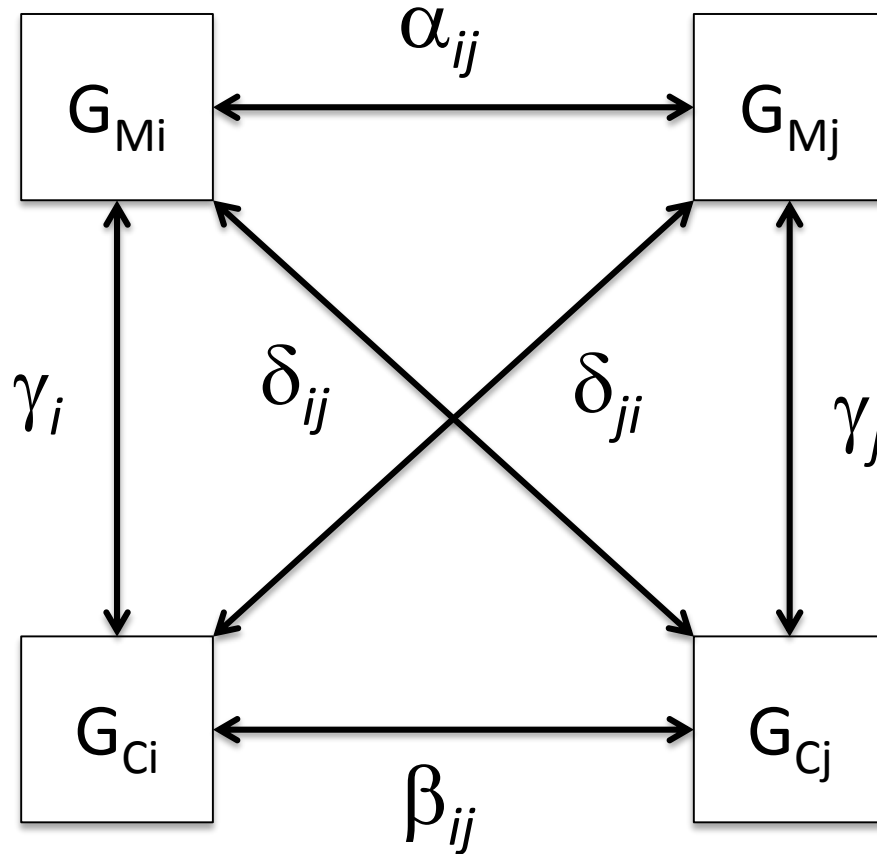
Q=Covariance arising from overlap between genes with direct and indirect effect (“passive rGE”, +ve or -ve) (M+Q->”C” in ACE model)

Pattern of Genetic Relatedness Within and Between Typical Unrelated Mother-Child Pairs

Pair i

Pair j

Mother



Offspring

$$E\gamma_i = 1/2$$

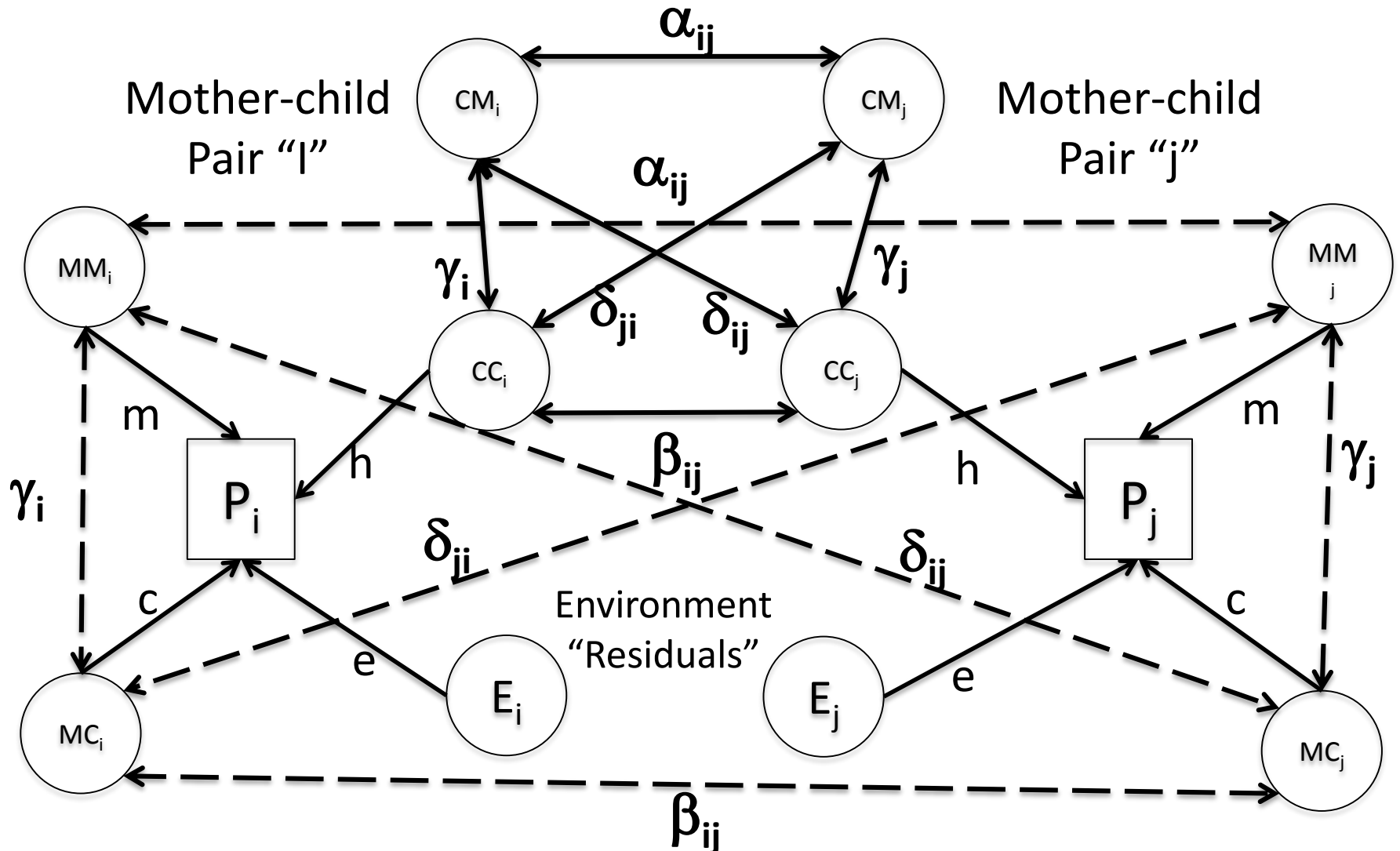
Components of genetic relatedness matrix for mothers and offspring

Relationship	Genetic component	Mothers		Offspring	
		MM	CM	MC	CC
Mothers	MM	A	0	D	0
	CM	0	A	0	D
Offspring	MC	D'	0	B	0
	CC	0	D'	0	B

Notes: Key to genetic components (c.f. Figure 2): MM=maternal copies of genes having indirect maternal effects (m) when present in the mother; CM = maternal copies of genes having direct effects (h) when present in offspring; MC= offspring copies of genes having indirect maternal effects (m) when present in the mother (may also have a direct effect, $c > 0$, on offspring when present in offspring); CC = offspring copies of genes having direct effects (h) when present in offspring. Component matrices are $N \times N$ where N =number of mother-child pairs in the sample.

Including Offspring Phenotype in Genetic Model

Genes with offspring-specific effects



Genes with maternal effects (may also have direct expression in child)

Covariance structure of unrelated individuals
given empirical estimates of genetic relatedness

$$\Sigma = \mathbf{A}\mathbf{M} + \mathbf{B}\mathbf{G} + \Delta\mathbf{Q} + \mathbf{I}\mathbf{E}$$

where $\Delta = \mathbf{D} + \mathbf{D}'$

Given phenotypes of N subjects and genetic relatedness matrices, **A**, **M** and **D** estimated from SNPs, can estimate genetic parameters by FIML (openMX) or REML (GCTA using Yang et als. program).

Can Fudge it in GCTA Package

Because Yang et al's GCTA program allows total covariance to be partitioned among multiple (independent) sources (chromosomes)

Thus:

$$\Sigma = \mathbf{B}_1 \mathbf{G}_1 + \mathbf{B}_2 \mathbf{G}_2 + \mathbf{B}_3 \mathbf{G}_3 + \dots + \mathbf{I}E$$

where G_i is the genetic variance contributed by the i th chromosome and B_i the genetic relatedness derived from SNPs on that chromosome. Our M-GCTA model has a similar linear form.

OpenMX vs GCTA for extended models

openMX

- (Currently) FIML. Impossibly slow with large N (>2000)
- Handles non-linear (path) models and constraints
- Need external programming to compute relatedness matrices
- Extends to multivariate models
- “Ours” – Easier to mess with

GCTA (Yang et al)

- REML very fast with large enough N (5000+)
- Unconstrained linear (variance components) model
- Computes relatedness matrices (rapidly)
- Limited extension to multi(bi-)variate case
- “Someone else’s” – Harder to mess with.

Outline of simulation

- Simulation done in R on MAC. Model fitted in OpenMx
- 1000 trios (mother, father, child); 500 SNPS.
- Decide which SNPs have direct or maternal effects, both or neither.
- Sample direct and maternal effects for each SNP.
- Simulate SNPs for parents and offspring trios
- Generate offspring phenotypes from maternal and offspring SNPs and residuals
- Compute genetic relatedness matrices (A,B and D) across all mother-child pairs (see Yang et al. for algebra).
- Recover ML estimates of G, M and Q using OpenMx
- Fit-reduced models and compare

Results of fitting “GCTA” model for direct and maternal effects to simulated offspring phenotypes in random samples of genotyped mother-child pairs (N=1000)

Model		μ	G	M	Q	E	-2lnl	χ^2	d.f.	P%
1	Simulated	100.29	25.12	0.	0.	3.86	-	-	-	-
	Full	100.19	24.23	-0.19	0.63	4.20	4904.71	-	-	-
	Q=0	100.19	24.93	-0.14	0.	4.16	4907.34	2.63	1	
	M=Q=0	100.19	24.91	0.	0.	4.05	4908.31	3.60	2	17
	G=Q=0	100.25	0.	7.23	0.	22.94	6152.00	1247.29	2	<10 ⁻⁶
	M=G=Q=0	100.29	0.	0.	0.	30.25	6247.31	1342.60	3	<10 ⁻⁶
2	Simulated	99.88	12.34	12.37	-0.14	3.86	-	-	-	-
	Full	99.79	11.74	12.28	1.05	4.02	5212.25	-	-	-
	Q=0	99.79	11.87	12.40	0.	4.00	5213.44	1.19	1	55
	M=Q=0	99.78	15.58	0.	0.	13.32	5797.36	575.11	2	<10 ⁻⁶
	G=Q=0	99.85	0.	16.43	0.	12.61	5759.37	547.12	2	<10 ⁻⁶
	M=G=Q=0	99.88	0.	0.	0.	29.07	6207.51	995.26	3	<10 ⁻⁶
3	Simulated	100.50	0.	25.74	0.	3.86	-	-	-	-
	Full	100.41	-0.20	24.66	0.57	4.14	4891.65	-	-	-
	Q=0	100.41	-0.17	25.27	0.	4.12	4893.93	2.28	1	13
	M=Q=0	100.49	7.81	0.	0.	23.03	6164.98	1273.33	2	<10 ⁻⁶
	G=Q=0	100.42	0.	25.21	0.	3.99	4895.57	3.92	2	14
	M=G=Q=0	100.50	0.	0.	0.	30.67	6261.20	1369.55	3	<10 ⁻⁶
4	Simulated	125.77	15.71	15.78	5.89	3.86	-	-	-	-
	Full	125.55	15.41	15.89	6.50	4.02	5284.08	-	-	-
	Q=0	125.56	16.04	16.43	0.	3.98	5314.57	30.49	1	<10 ⁻⁶
	M=Q=0	125.62	25.87	0.	0.	16.38	6054.66	770.58	2	<10 ⁻⁶
	G=Q=0	125.63	0.	25.94	0.	15.44	6005.47	729.39	2	<10 ⁻⁶
	M=G=Q=0	125.77	0.	0.	0.	42.45	6586.25	1302.17	3	<10 ⁻⁶

Comparison of estimates obtained from simulated data by FIML in OpenMx and REML in GCTA.

Open Mx Results									
	m	G	M	Q	E	-2(L+C ₁)	χ^2	d.f.	P%
Simulated	125.77	15.71	15.78	5.89	3.86	-	-	-	-
Full	125.55	15.41	15.89	6.50	4.02	5284.08	-	-	-
Q=0	125.56	16.04	16.43	0.	3.98	5314.57	30.49	1	<10 ⁻⁶
M=Q=0	125.62	25.87	0.	0.	16.38	6054.66	770.58	2	<10 ⁻⁶
G=Q=0	125.63	0.	25.94	0.	15.44	6005.47	729.39	2	<10 ⁻⁶
M=G=Q=0	125.77	0.	0.	0.	42.45	6586.25	1302.17	3	<10 ⁻⁶

GCTA Results									
	m	G	M	Q	E	-2(L+C ₂)	c ²	d.f.	P%
Simulated	125.77	15.71	15.78	5.89	3.86	-	-	-	-
Full	125.55	15.41	15.89	6.50	4.02	3451.22	-	-	-
Q=0	125.56	16.04	16.43	0.	3.99	3481.28	30.66	1	<10 ⁻⁶
M=Q=0	125.62	25.87	0.	0.	16.40	4220.68	769.46	2	<10 ⁻⁶
G=Q=0	125.63	0.	25.94	0.	15.46	4171.56	720.34	2	<10 ⁻⁶
M=G=Q=0	NA	0.	0.	0.	42.49	4751.54	1300.32	3	<10 ⁻⁶

Application to Real Data

The Avon Longitudinal Study of Parents and Children ("ALSPAC")

UK Medical Research Council
Wellcome Trust
University of Bristol

[http://www.bris.ac.uk/alspac/researchers/data-
access/data-dictionary](http://www.bris.ac.uk/alspac/researchers/data-access/data-dictionary)

ALSPAC

- Early 1990's
- Avon County, UK
- Population-based
- Birth cohort study
- Longitudinal follow-up
- 14,541 women and children
- Mothers genotyped on the Illumina 660K SNP chip.
- Children genotyped on the Illumina 550K SNP chip.
- After cleaning etc. genetic relatedness computed for 4761 unrelated mother-offspring pairs

ALSPAC

Boyd A, Golding J, Macleod J, Lawlor DA, Fraser A, Henderson J, Molloy L, Ness A, Ring S, Davey Smith G (2013) Cohort Profile: the 'children of the 90s'--the index offspring of the Avon Longitudinal Study of Parents and Children. *Int J Epidemiol.* 42(1):111-27.

Fraser A, Macdonald-Wallis C, Tilling K, Boyd A, Golding J, Davey Smith G, Henderson J, Macleod J, Molloy L, Ness A, Ring S, Nelson SM, Lawlor DA (2013) Cohort Profile: the Avon Longitudinal Study of Parents and Children: ALSPAC mothers cohort. *Int J Epidemiol.* 42(1):97-110.

Proof of Principle

- Maternal, self-reported, pre-pregnancy BMI - should be no effects of offspring genotype.
- Neonatal birth length (crown-heel) – might show effects of both maternal and fetal genotype

Results of fitting “GCTA” model for direct and maternal effects to maternal BMI and birth length data in the ALSPAC cohort. Results are presented as standardized variance components.

Maternal BMI	Model	G	M	Q	E	-2lnl	χ^2	d.f.	P%
	Full	0.06 (0.10)	0.12 (0.11)	0.11 (0.08)	0.72 (0.12)	3813.926	-	-	-
	Q=0	0.13 (0.08)	0.20 (0.08)	-	0.66 (0.11)	3815.582	1.66	1	20
	M=Q=0	0.19 (0.08)	-	-	0.80 (0.09)	3821.224	7.30	2	2.6
	G=Q=0	-	0.25 (0.09)	-	0.75 (0.09)	3817.952	4.03	2	13
	M=G=Q=0	-	-	1		3826.818	12.89	3	5x10 ⁻³
Birth Length									
	Full	0.11 (0.12)	0.14 (0.12)	0.07 (0.09)	0.68 (0.13)	3592.748	-	-	-
	Q=0	0.16 (0.10)	0.20 (0.10)	-	0.64 (0.12)	3593.402	0.65	1	42
	M=Q=0	0.21 (0.09)	-	-	0.79 (0.10)	3597.240	4.49	2	11
	G=Q=0	-	0.24 (0.10)	-	0.76 (0.10)	3596.408	3.66	2	16
	M=G=Q=0	-	-	1		3602.390	9.64	3	2.2

Conclusions?

- It seems to work – so far
- Subject to all the ifs and buts that go with using genome-wide SNPs to estimate genetic relatedness

Next Steps

- Extend to other dyadic relationships: other relatives, social networks (peers etc.)
- Multivariate systems
- Causal models and intermediate pathways
- Implications of assortative mating?
- Non-additive genetic effects
- Get it to work properly in OpenMx to capitalize on flexible modeling capabilities (MIKE et al.!!!!!!)