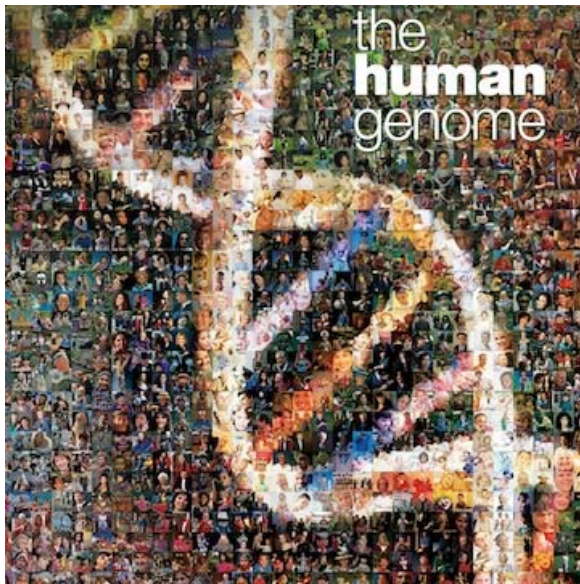


Genetic epidemiology in the genomic age:
Missing Heritability &
The Role of Twin Studies in the Genomic Era

Michel Nivard, Conor Dolan & Nick Martin

The genomic era: Mapping the genome



2001 Draft of the human genome released, 2003 final sequenced genome released.

The final genome contained ~ 3 Billion base pairs

The genomic era: Mapping the genome



The genomic era: Mapping genomic variation

2005-10-24: **HapMap Public Release #19**
Genome sequence for 270 individuals
4 populations

Human genome project: GCTATCGATCACT

HAPMAP:

GCTATCG**A**TCACT

GCTATCG**G**TCACT

GCTATCG**A**TCACT

GCTATCG**G**TCACT

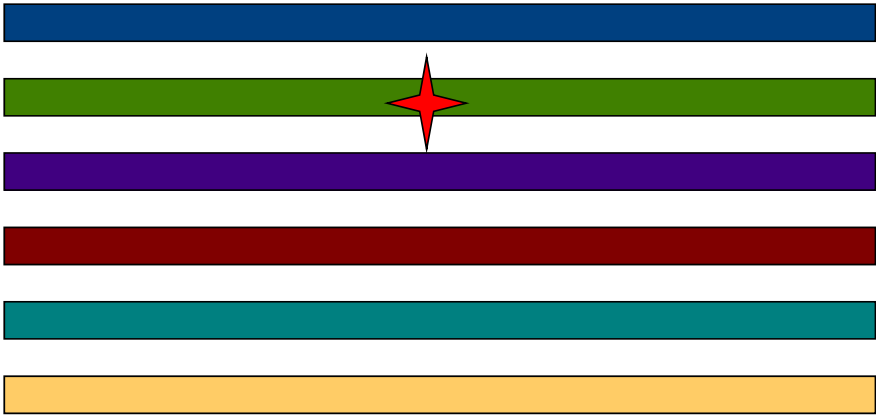


The genomic era: genotyping tag SNPs

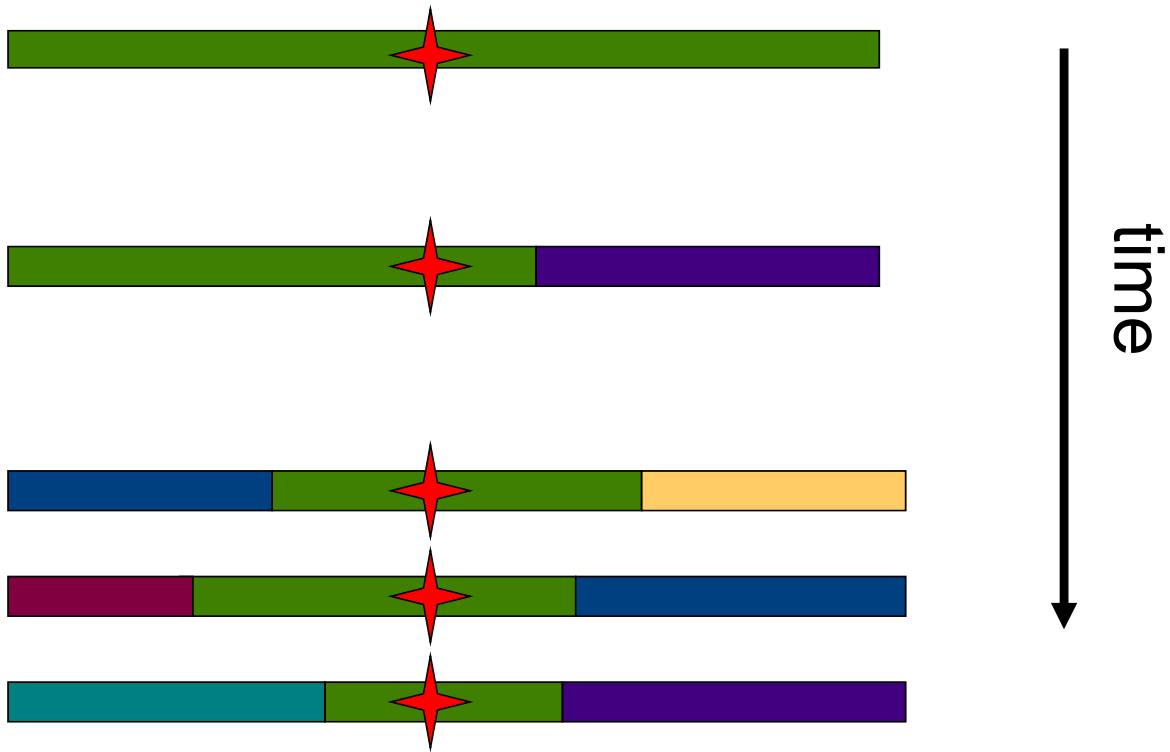
how many angels on the head of a pin? I don't know



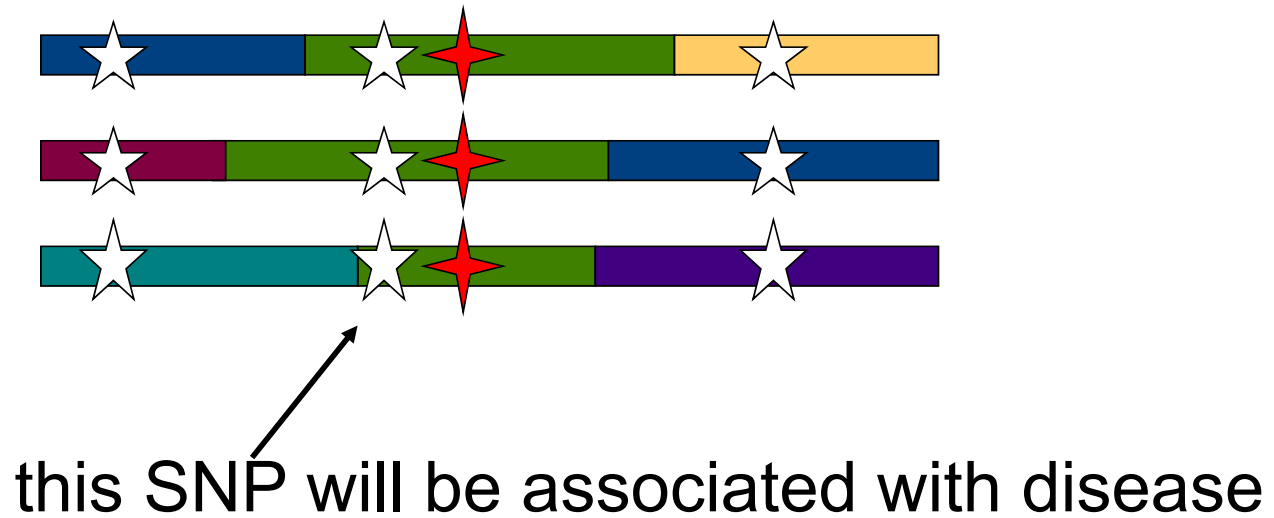
The genomic era: Mapping genomic variation



The genomic era: Mapping genomic variation



The genomic era: Mapping genomic variation



The genomic era: Case control studies



Allele **G** is 'associated' with disease

The genomic era: Case control studies

- Each individual contributes two counts to 2x2 table.
- Test of association

$$X^2 = \sum_{i=0,1} \sum_{j=A,U} \frac{(n_{ij} - E[n_{ij}])^2}{E[n_{ij}]}$$

where $E[n_{ij}] = \frac{n_{i.} \cdot n_{.j}}{n_{..}}$

	Cases	Controls	Total
G	n_{1A}	n_{1U}	$n_{1.}$
T	n_{0A}	n_{0U}	$n_{0.}$
Total	$n_{.A}$	$n_{.U}$	$n_{..}$

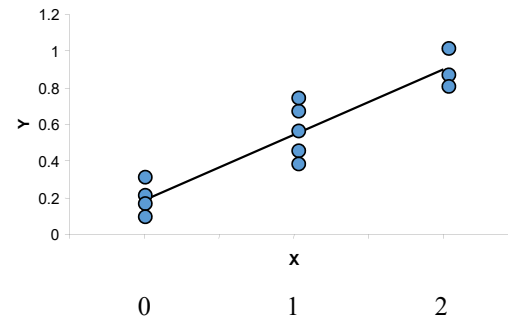
- X^2 has χ^2 distribution with 1 degrees of freedom under null hypothesis.

The genomic era: continuous phenotype

$$Y_i = \alpha + \beta X_i + e_i$$

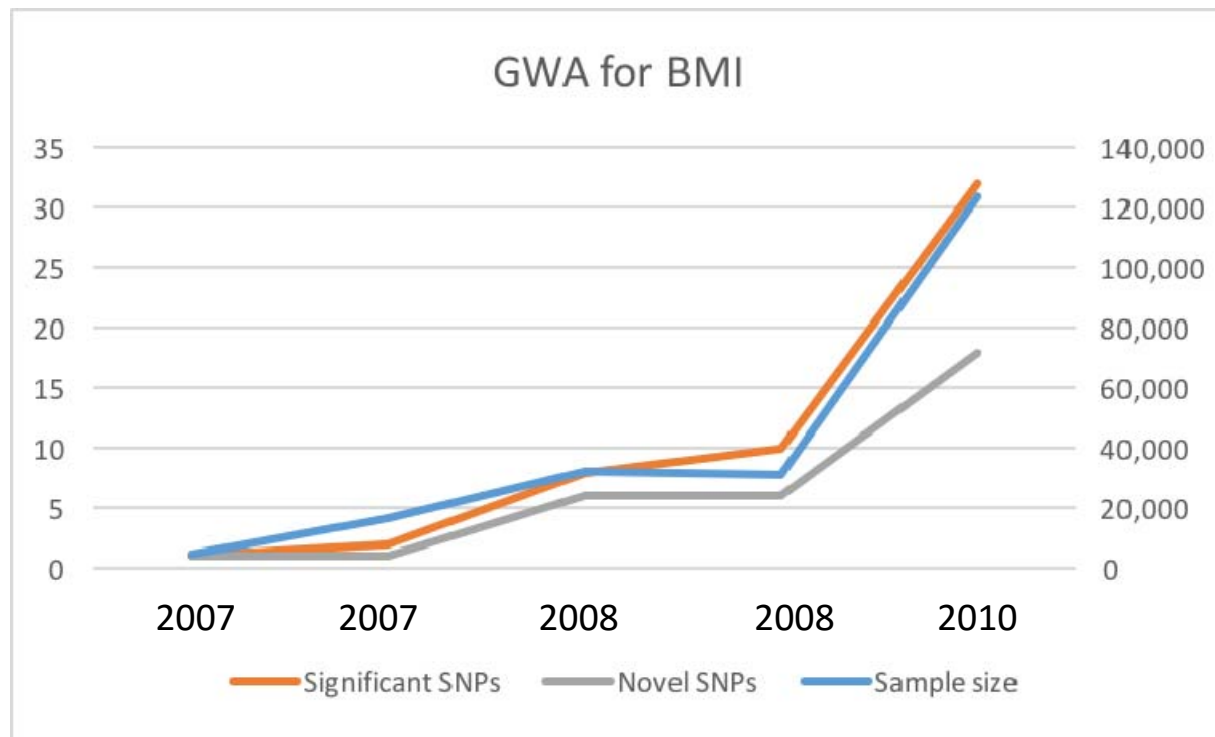
where

$Y_i =$ trait value for individual i
 $X_i =$ number of 'A' alleles an individual has

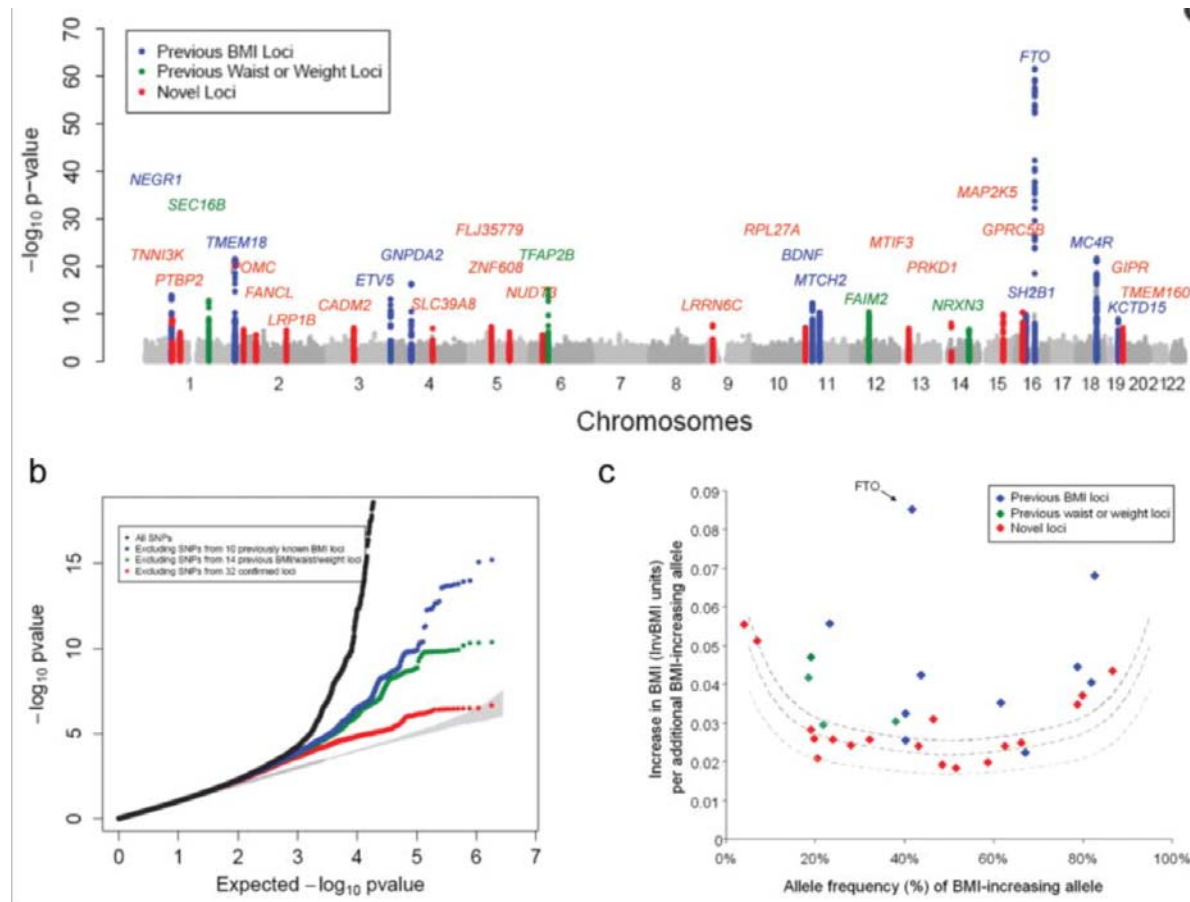


Association test is whether $\beta \neq 0$

The genomic era: BMI



The genomic era: BMI



The genomic era: “Missing” heritability



The case of the missing heritability

B. Maher The case of the missing heritability
Nature 2008

The genomic era: “Missing” heritability

The variance explained by genotyped loci related to complex traits, is substantially smaller than the additive genetic variance estimated in twin studies of the same traits

$$h_{missing}^2 = 1 - \left(\frac{R_{snp}^2}{h^2}\right)$$

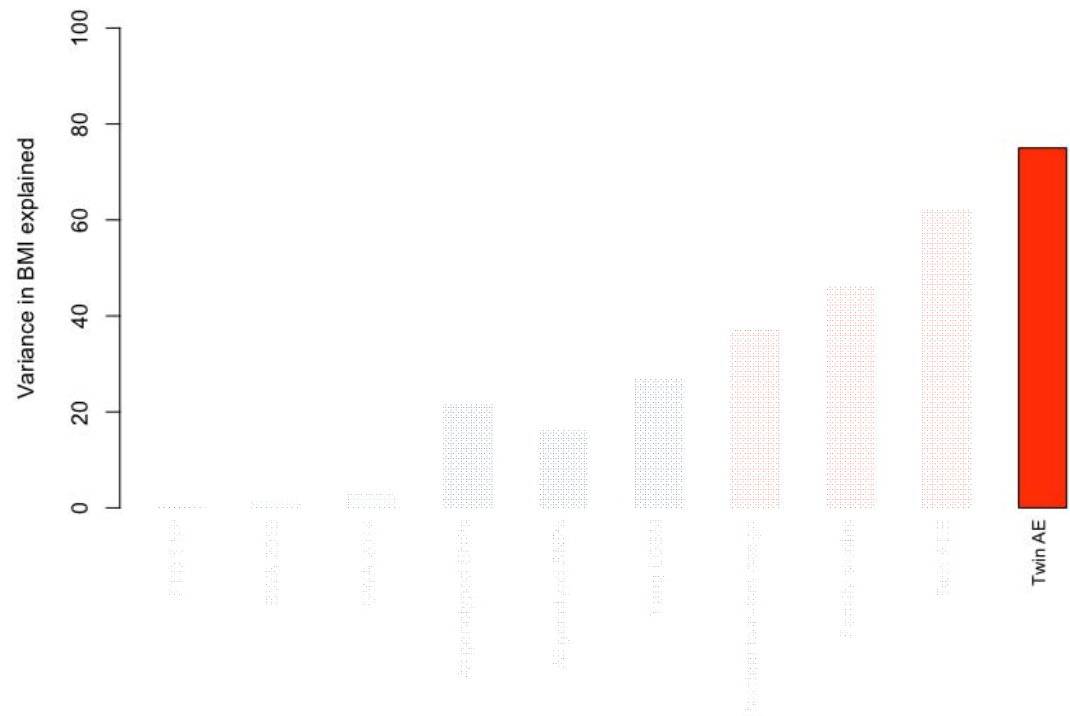


The case of the missing heritability

B. Maher The case of the missing heritability
Nature 2008

“Missing” heritability: BMI

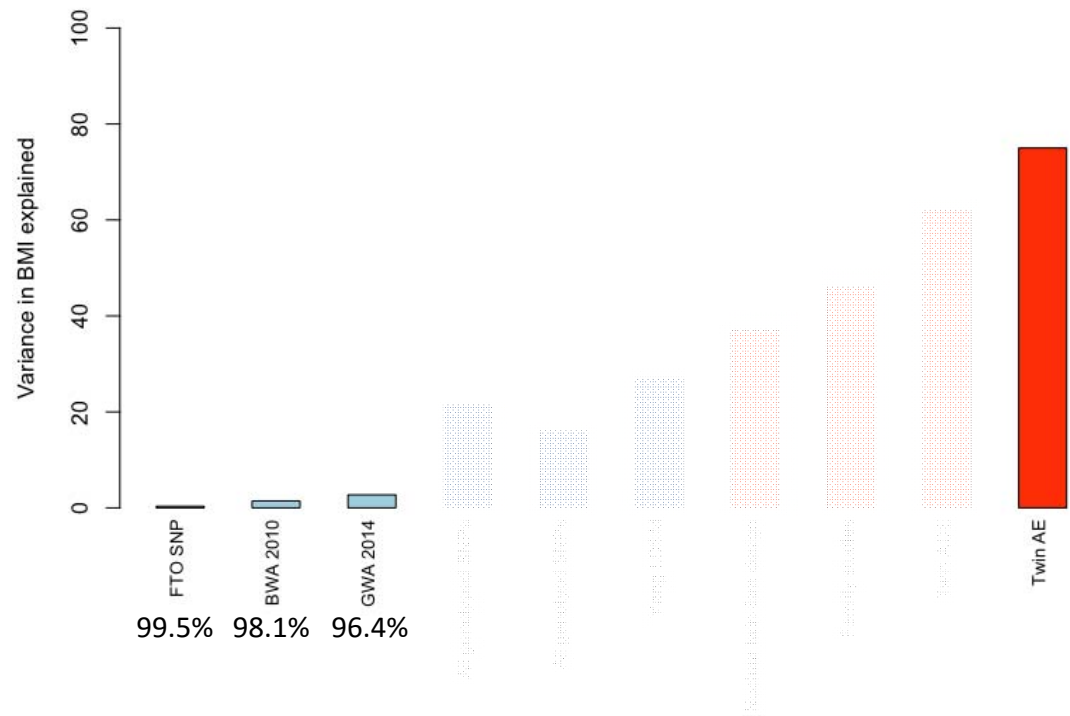
Elks et al (2012) Meta analysis of twin studies:
Variance in BMI explained by additive genetic effects:
75% (74%-76%)



“Missing” heritability: BMI

Elks et al (2012) Meta analysis of twin studies:
Variance in BMI explained by additive genetic effects:
75% (74%-76%)

Speliotes et al GWA BMI 2010 (n= 123,865):
Strongest locus: **0.34%** of variance in BMI
All significant loci(32): **1.45%** of variance in BMI

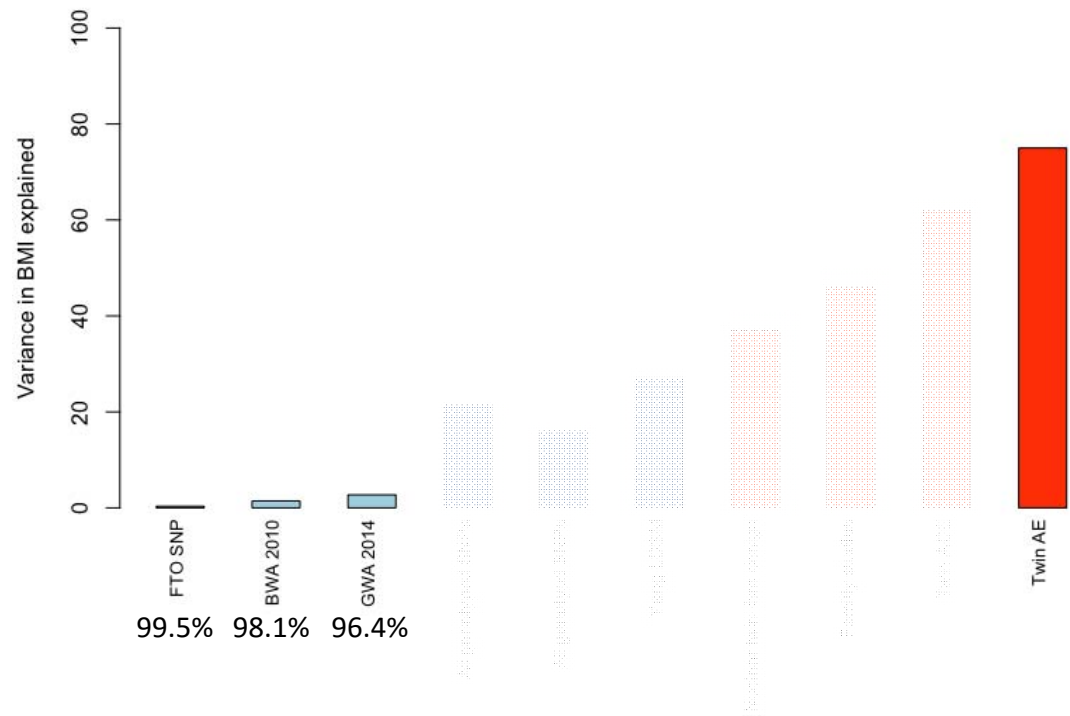


“Missing” heritability: BMI

Elks et al (2012) Meta analysis of twin studies:
Variance in BMI explained by additive genetic effects:
75% (74%-76%)

Speliotes et al GWA BMI 2010 (n= 123,865):
Strongest locus: **0.34%** of variance in BMI
All significant loci(32): **1.45%** of variance in BMI

Locke et al GWA BMI 2015 (n= 339,224):
All significant loci(97): **2.7%** of variance in BMI

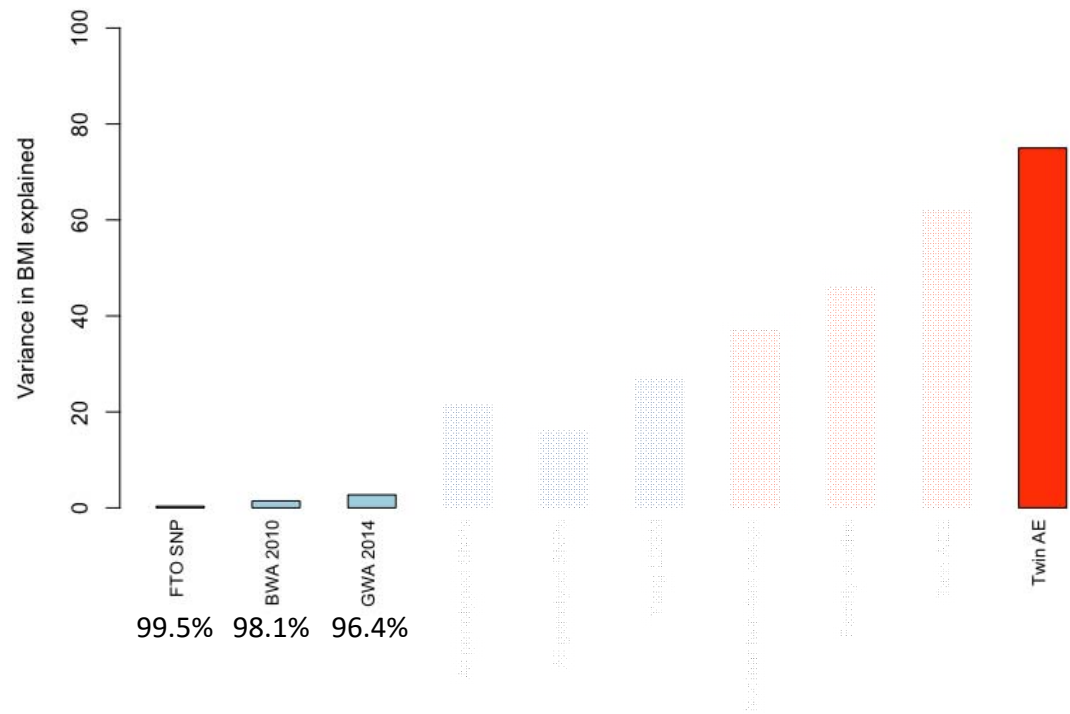


“Missing” heritability: BMI

Elks et al (2012) Meta analysis of twin studies:
Variance in BMI explained by additive genetic effects:
75% (74%-76%)

Speliotes et al GWA BMI 2010 (n= 123,865):
Strongest locus: **0.34%** of variance in BMI
All significant loci(32): **1.45%** of variance in BMI

Locke et al GWA BMI 2015 (n= 339,224):
All significant loci(97): **2.7%** of variance in BMI
All loci: **2.7%** of variance in BMI

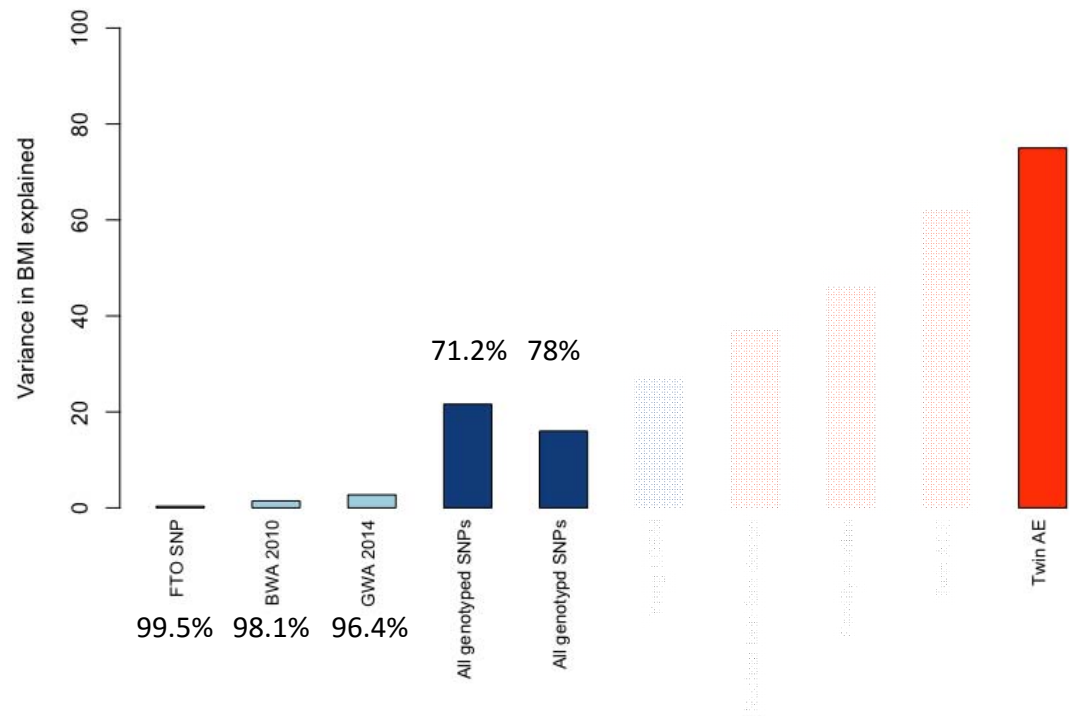


“Missing” heritability: BMI

Elks et al (2012) Meta analysis of twin studies:
Variance in BMI explained by additive genetic effects:
75% (74%-76%)

Yang et al (2012) GCTA:
16.5% of variation in BMI explained by genotyped SNPs

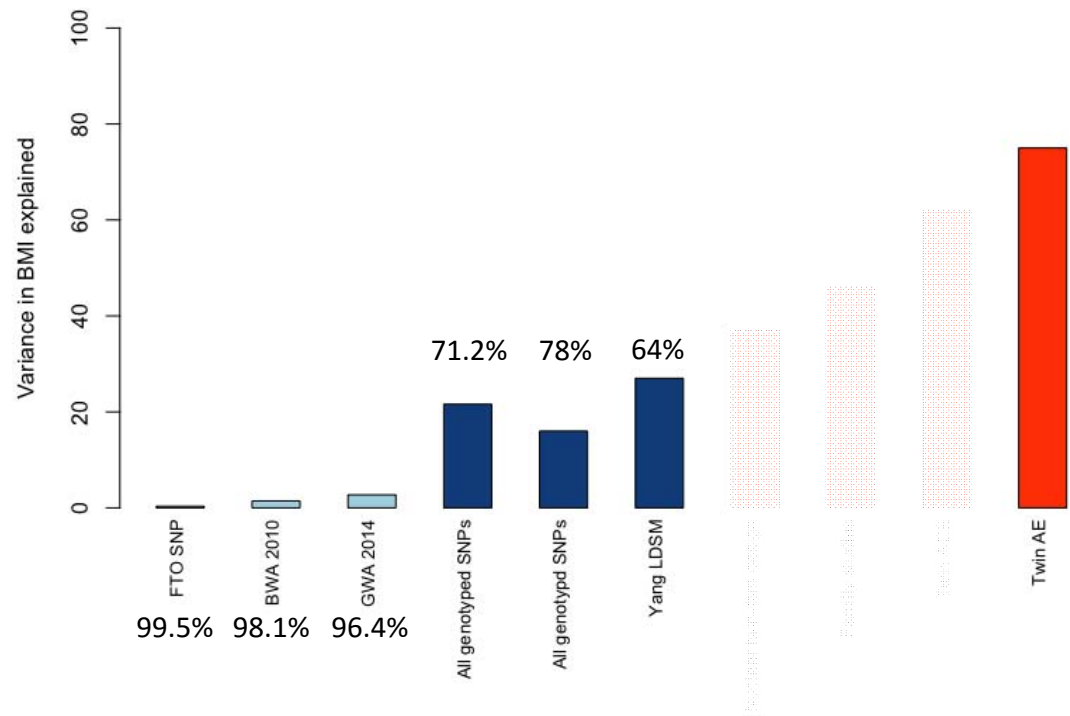
Lock et al (2015) GCTA:
21.6% of variation in BMI explained by measured SNPs



“Missing” heritability: BMI

Elks et al (2012) Meta analysis of twin studies:
Variance in BMI explained by additive genetic effects:
75% (74%-76%)

Yang et al (2014) GCTA-LDMS:
27% of variance explained by imputed SNPs



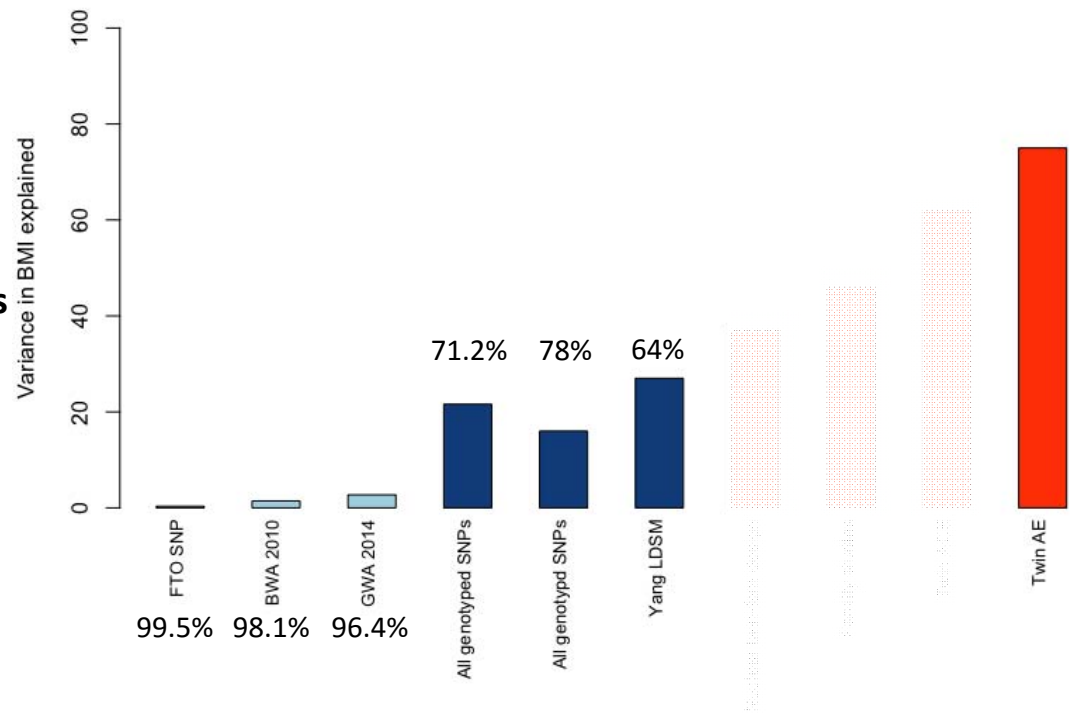
“Missing” heritability: BMI

Elks et al (2012) Meta analysis of twin studies:
Variance in BMI explained by additive genetic effects:
75% (74%-76%)

Yang et al (2014) GCTA-LDMS:
27% of variance explained by imputed SNPs

Hypothesized causes

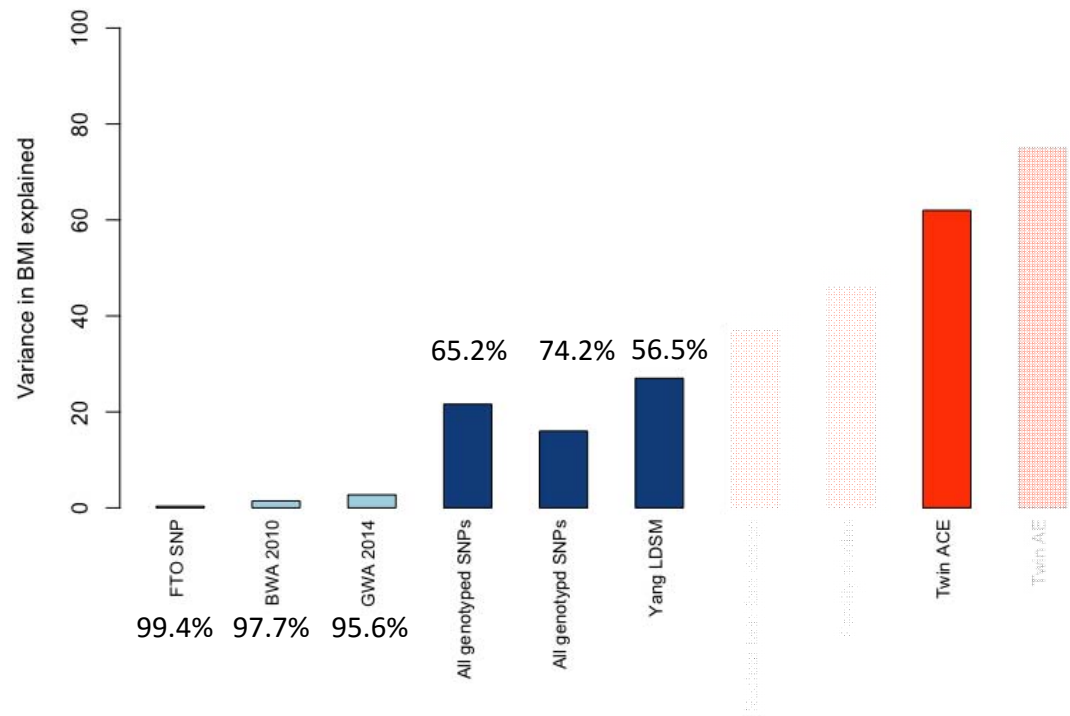
- a) Rare genetic variation
- b) Overestimation of additive variance in twin models**



“Missing” heritability: BMI

Hypothesized causes

1. Rare genetic variation
2. **Overestimation of additive variance in twin model**
 - a) Allow for common environment
 - a) h^2 75% \rightarrow 62%



“Missing” heritability: BMI

Hypothesized causes

1. Rare genetic variation

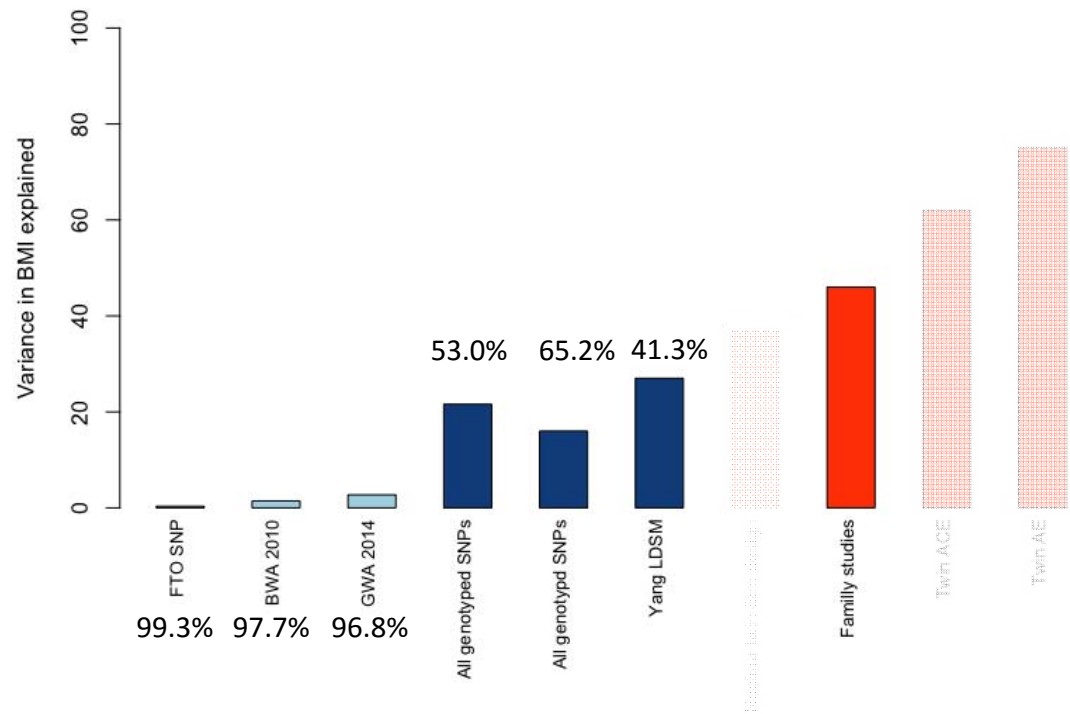
2. **Overestimation of additive variance in twin model**

a) Allow for common environment

a) h^2 75% \rightarrow 62%

b) **Rely on family members instead of twins**

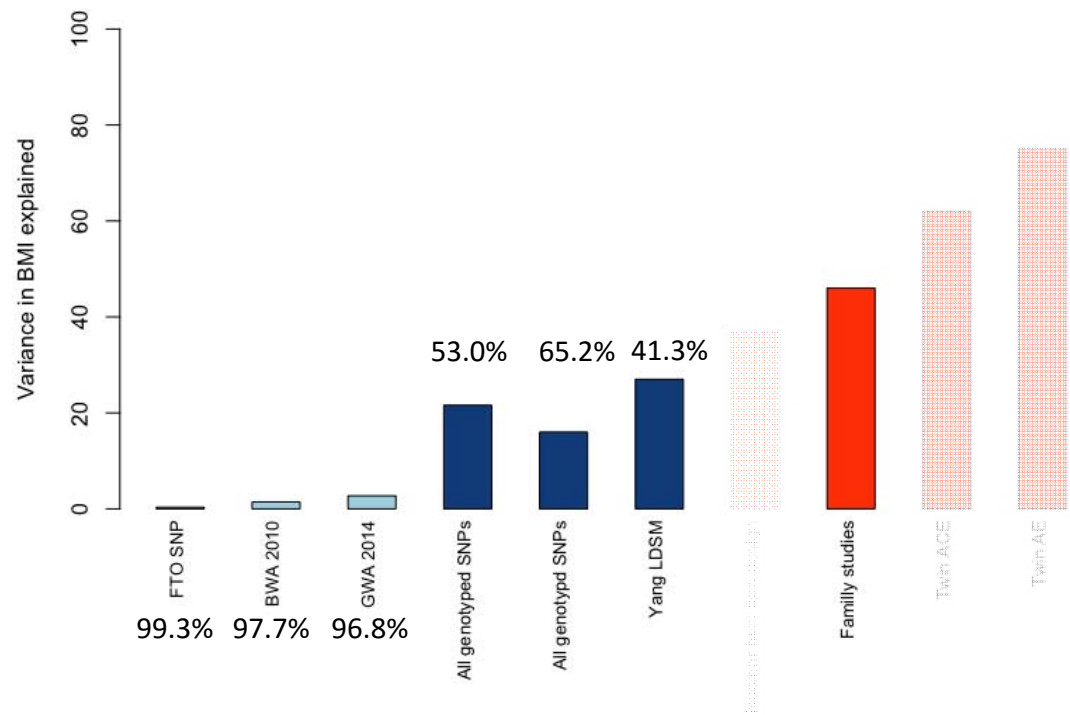
a) h^2 75% \rightarrow 46%



“Missing” heritability: BMI

Hypothesized causes

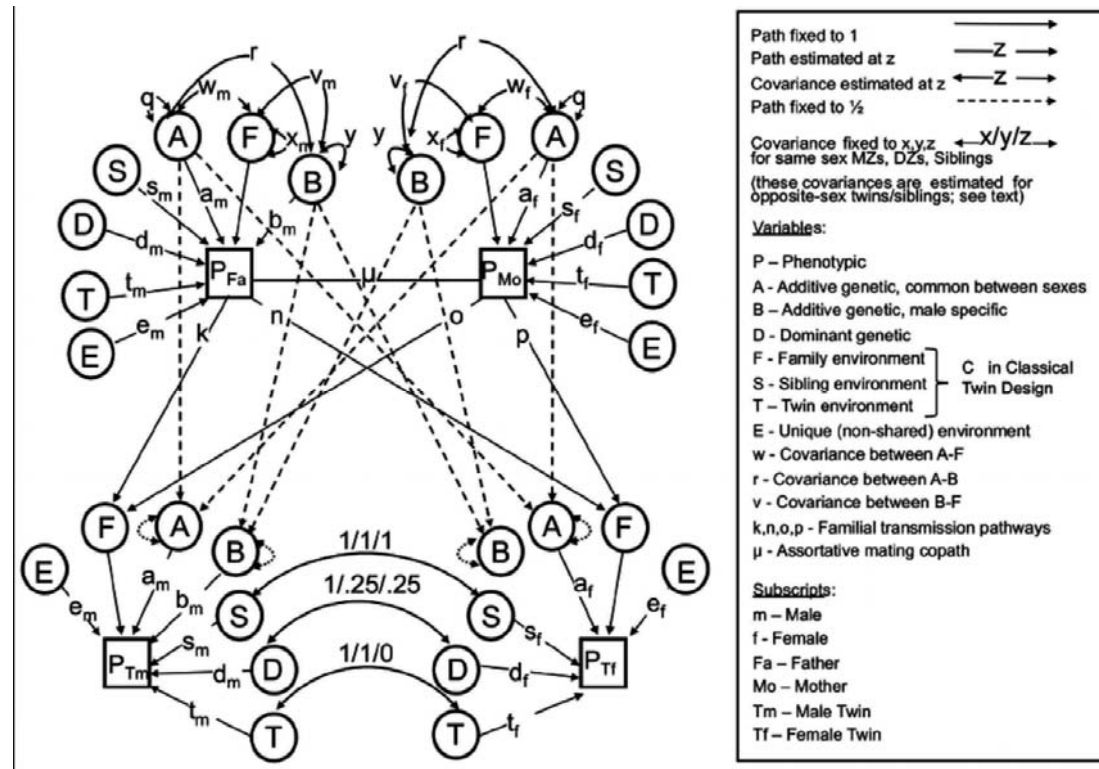
1. Rare genetic variation
2. **Overestimation of additive variance in twin model**
 - a) Allow for common environment
 - a) h^2 75% \rightarrow 62%
 - b) **Rely on family members instead of twins**
 - a) h^2 75% \rightarrow 46%
 - c) **Fit the Nuclear Twin-Family model**



“Missing” heritability: BMI

Hypothesized causes

1. Rare genetic variation
2. **Overestimation of additive variance in twin model**
 - a) Allow for common environment
 - a) h^2 75% -> 62%
 - b) Rely on family members instead of twins
 - a) h^2 75% -> 46%
 - c) Fit the Nuclear Twin-Family model
 - a) Non-random mating
 - b) Dominance variation
 - c) Common environment



Maes, Eaves, Neale 1997

“Missing” heritability: BMI

Hypothesized causes

1. Rare genetic variation

2. **Overestimation of additive variance in twin model**

a) Allow for common environment

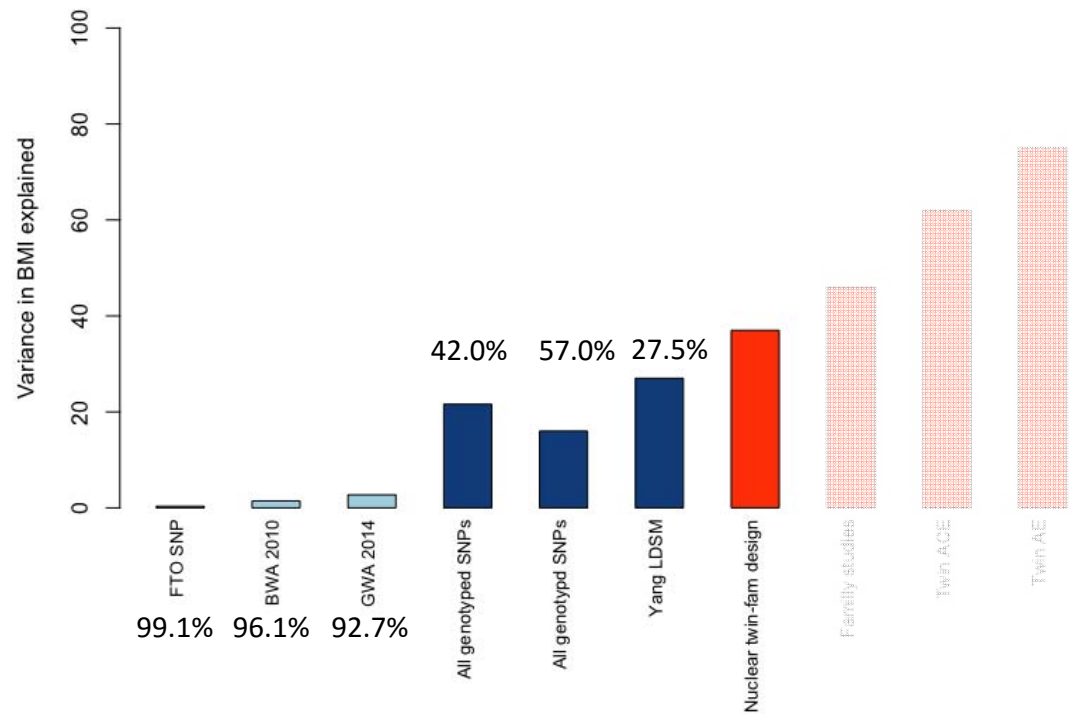
a) h^2 75% \rightarrow 62%

b) Rely on family members instead of twins

a) h^2 75% \rightarrow 46%

c) **Fit the Nuclear Twin-Family model**

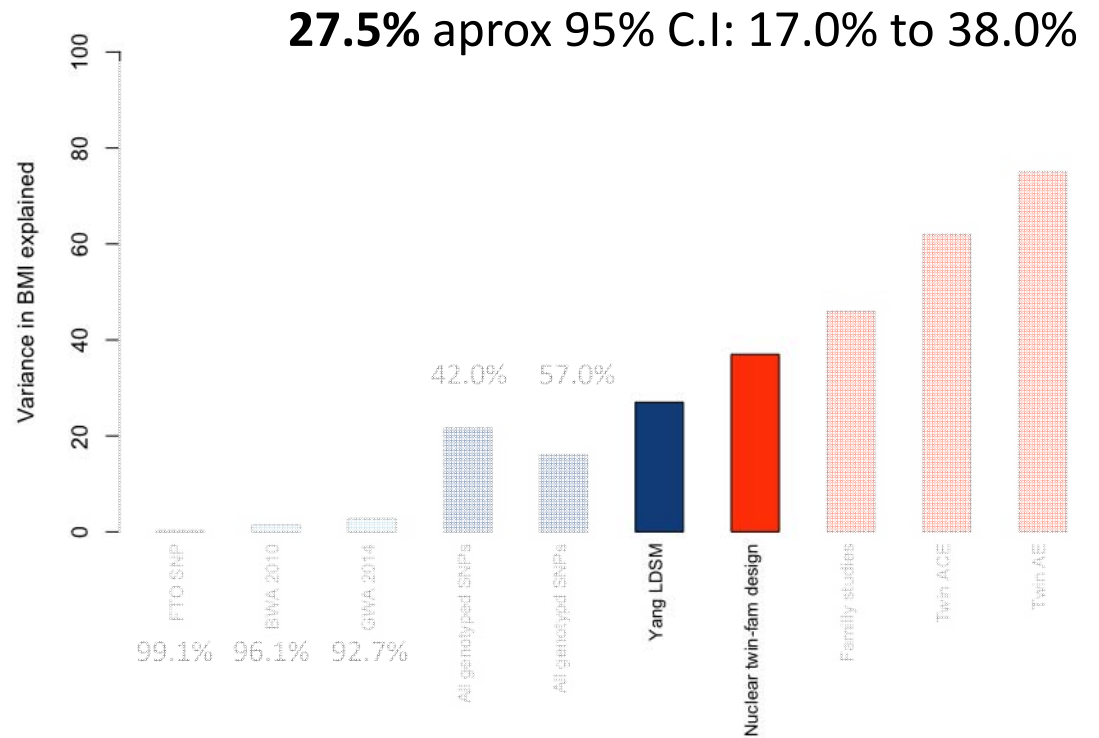
a) h^2 75% \rightarrow ~37%



“Missing” heritability: BMI

Hypothesized causes

1. Rare genetic variation
2. **Overestimation of additive variance in twin model**
 - a) Allow for common environment
 - a) h^2 75% \rightarrow 62%
 - b) Rely on family members instead of twins
 - a) h^2 75% \rightarrow 46%
 - c) **Fit the Nuclear Twin-Family model**
 - a) h^2 75% \rightarrow ~37%



“Missing” heritability: BMI

27.5% aprox C.I: 17.0% to 38.0%

The ratio of the “ Best” estimate based on genotyped distantly related individuals over the “ Best” estimate from a complex twin-family

Twin model captures:

Additive genetic variation

Dominance variation

Common environment “Special twin environment”

Assortment

Twin model assumes:

Absence of GxAge interaction

GCTA captures:

99% of common additive genetic effects, 68% of additive rare genetic effects*

GCTA assumes:

No assortment

* Yang et al 2014

The (gen)omic era: the continued relevance of twin studies.

REVIEWS

STUDY DESIGNS

The continuing value of twin studies in the omics era

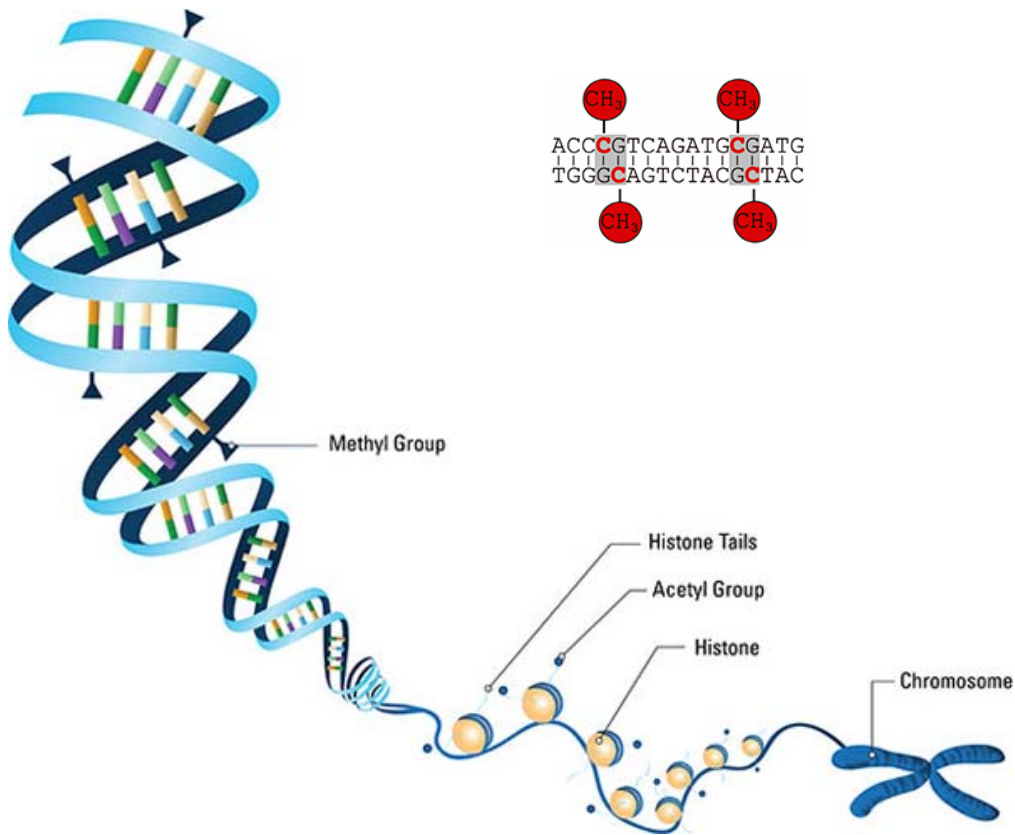
Jenny van Dongen¹, P. Eline Slagboom², Harmen H. M. Draisma¹, Nicholas G. Martin⁵ and Dorret I. Boomsma¹

Abstract | The classical twin study has been a powerful heuristic in biomedical, psychiatric and behavioural research for decades. Twin registries worldwide have collected biological material and longitudinal phenotypic data on tens of thousands of twins, providing a valuable resource for studying complex phenotypes and their underlying biology. In this Review, we consider the continuing value of twin studies in the current era of molecular genetic studies. We conclude that classical twin methods combined with novel technologies represent a powerful approach towards identifying and understanding the molecular pathways that underlie complex traits.

Van Dongen, J., Slagboom, P. E., Draisma, H. H., Martin, N. G., & Boomsma, D. I. (2012). The continuing value of twin studies in the omics era. *Nature Reviews Genetics*, 13(9), 640-653.

Epigenetic variation

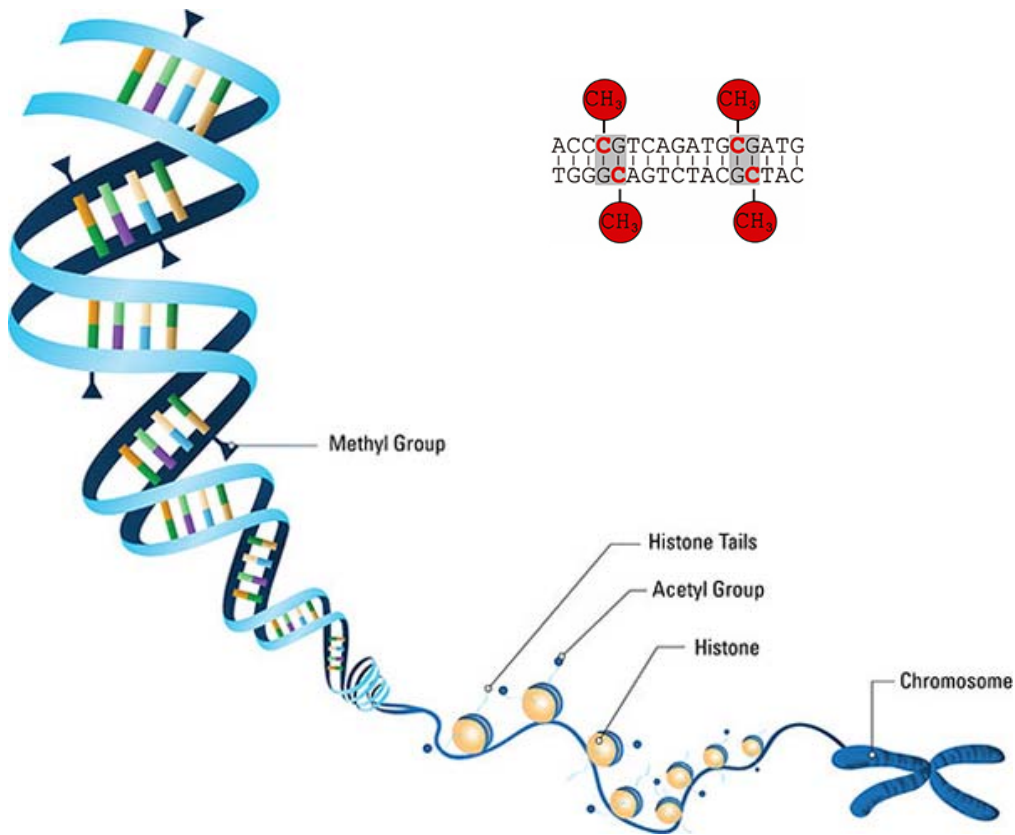
CpG methylation can vary across the lifespan



Epigenetic variation

CpG methylation can vary across the lifespan

CpG methylation can vary across tissues

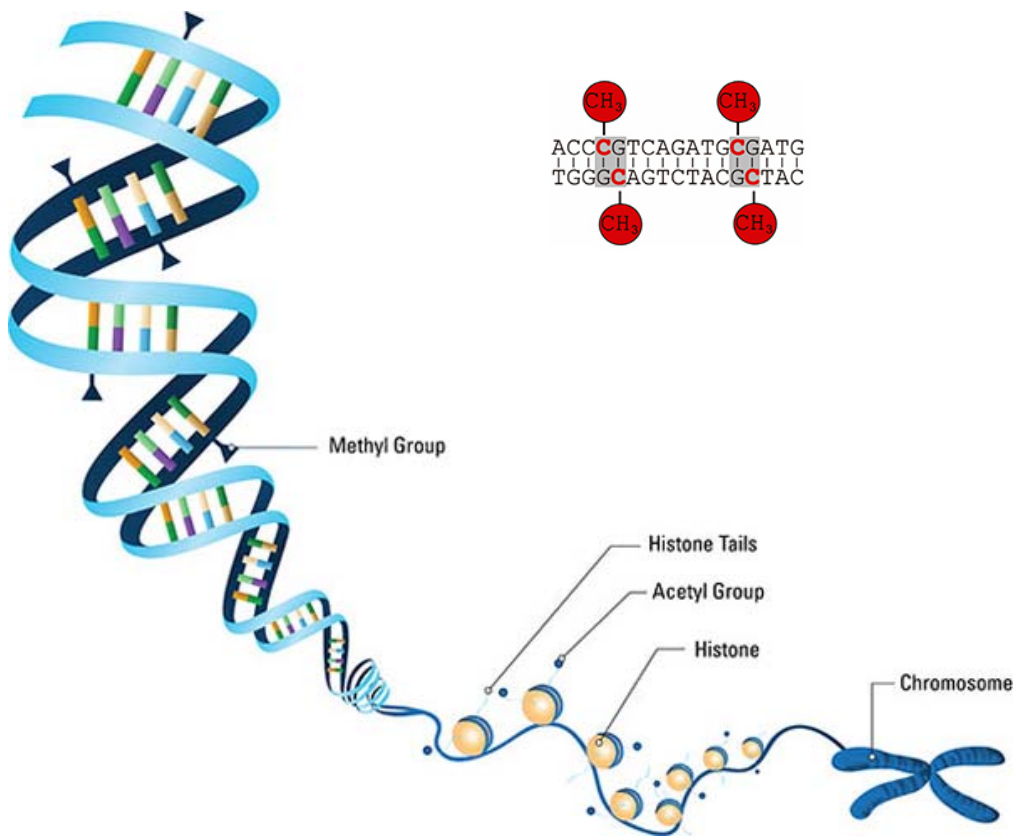


Epigenetic variation

CpG methylation can vary across the lifespan

CpG methylation can vary across tissues

CpG methylation changes the expression of genes



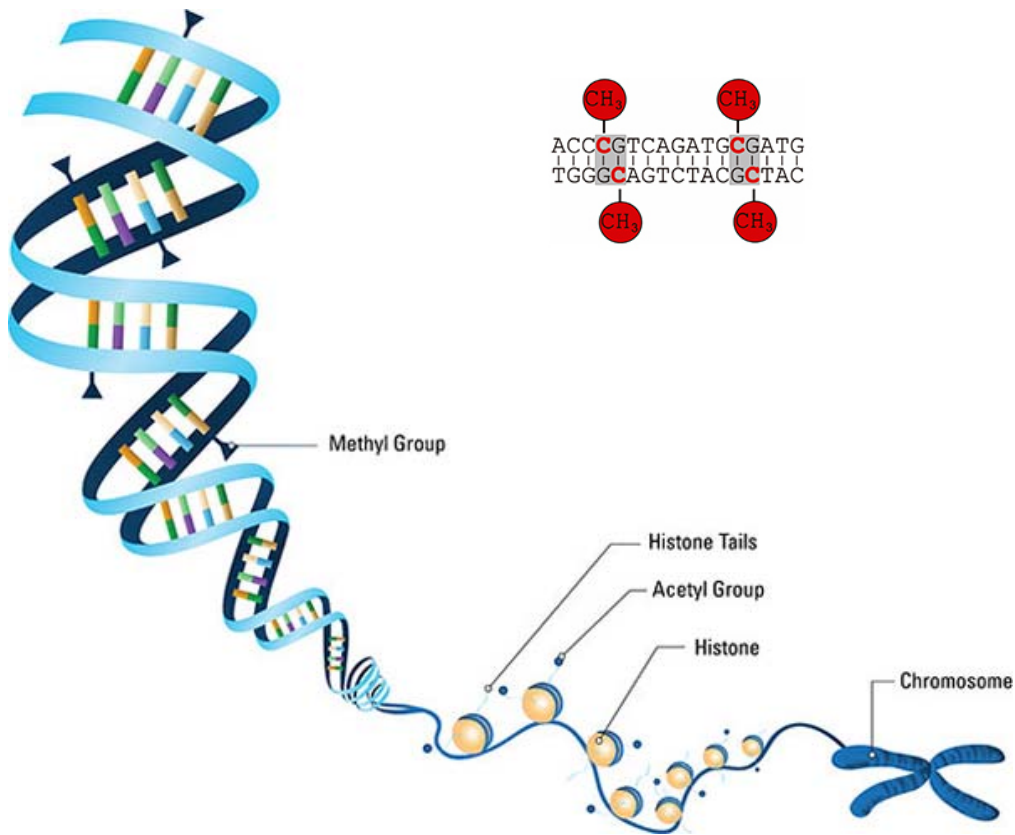
Epigenetic variation

CpG methylation can vary across the lifespan

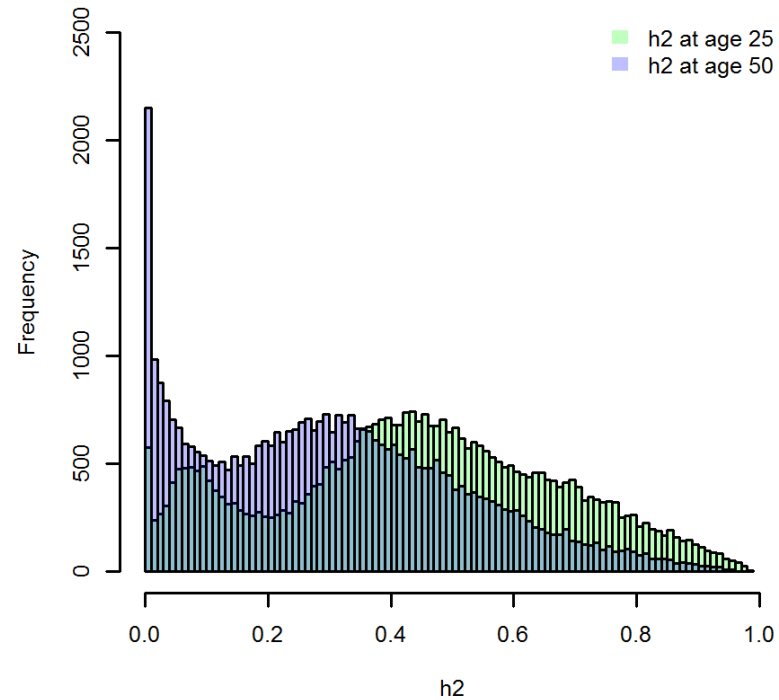
CpG methylation can vary across tissues

CpG methylation changes the expression of genes

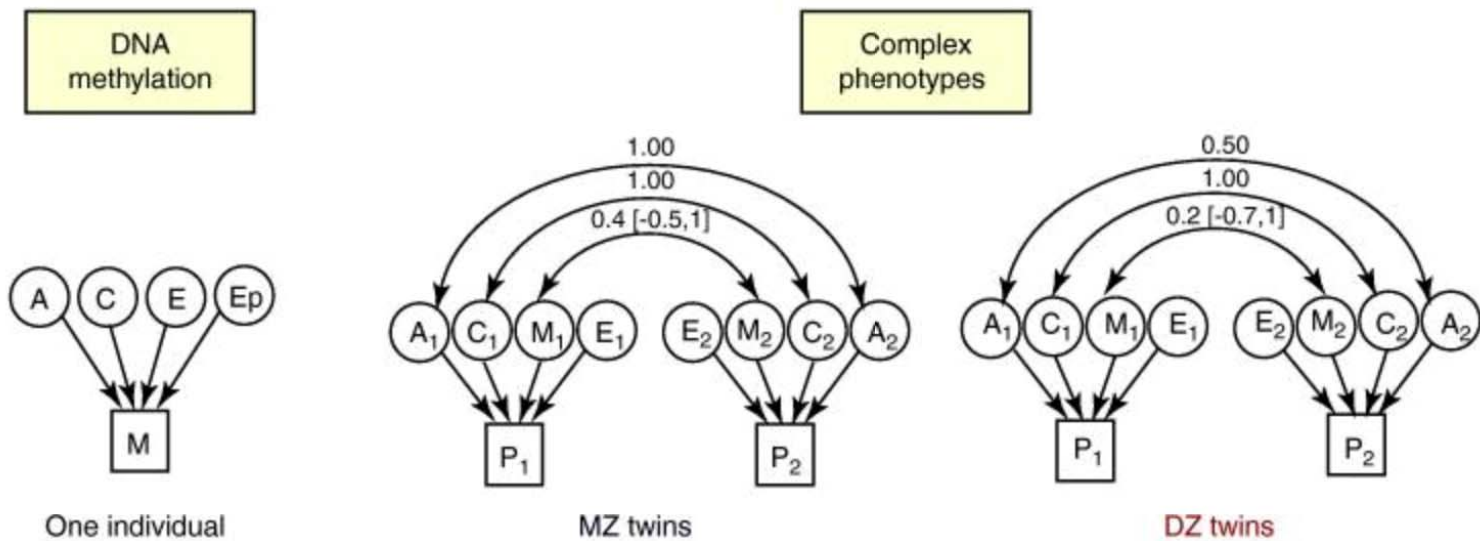
CpG Methylation is **NOT** identical in MZ twins, and does **NOT** correlate .5 between DZ twins



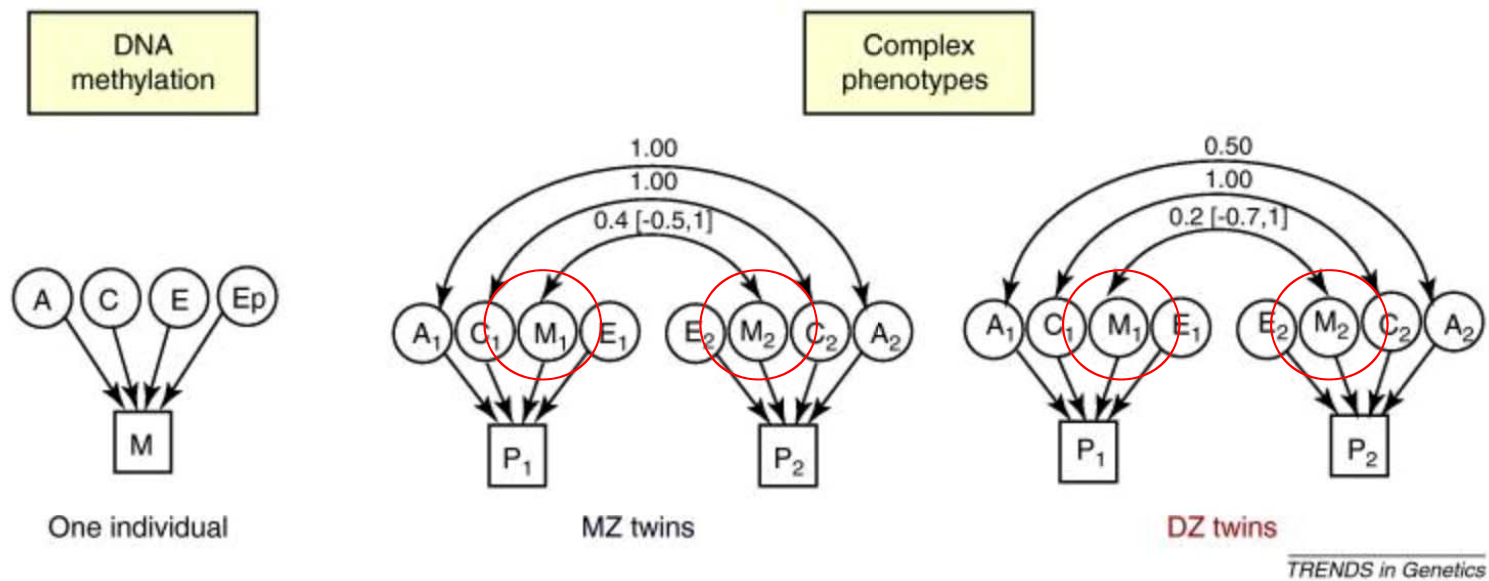
Heritability of CpG methylation



CpG Methylation in the twin model

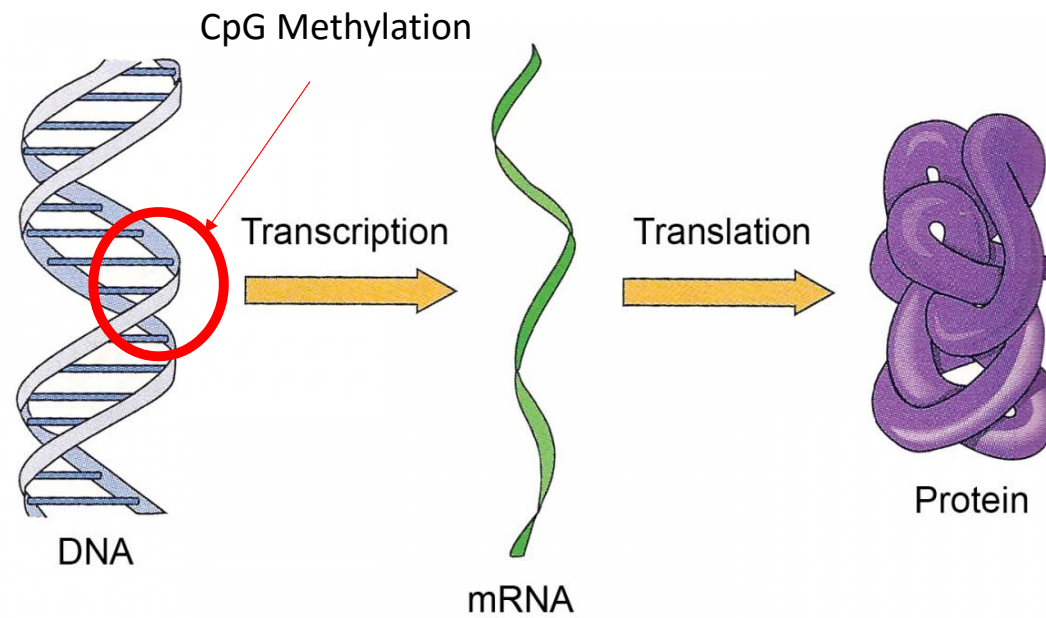


CpG Methylation in the twin model

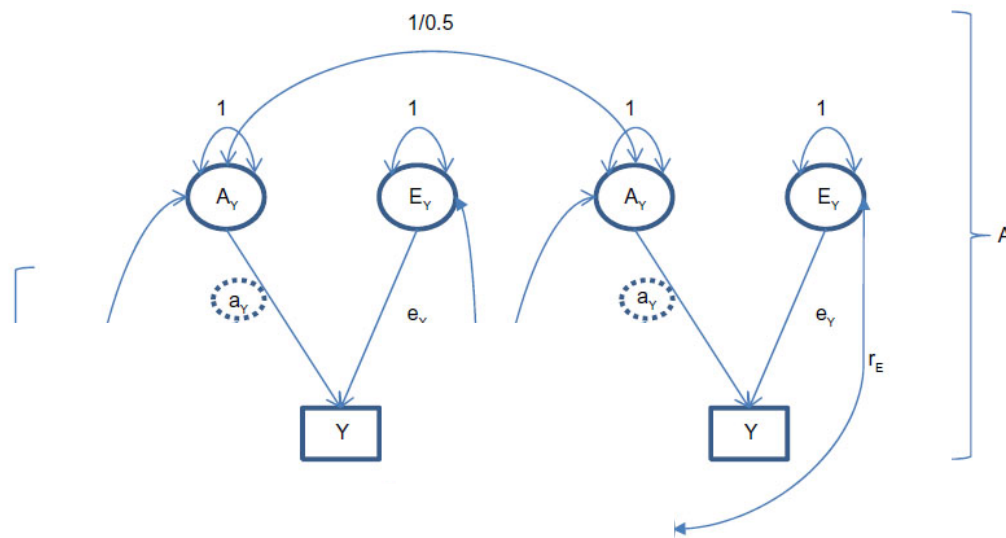


Bell & Spector (2011)

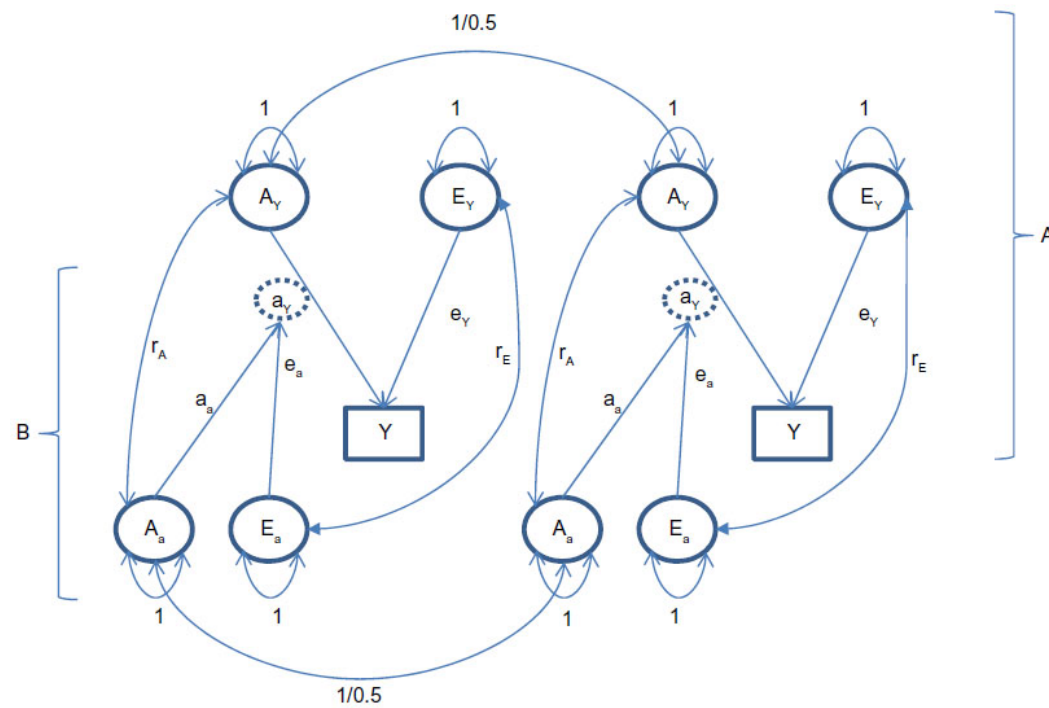
CpG Methylation in the twin model



CpG Methylation in the twin model



CpG Methylation in the twin model



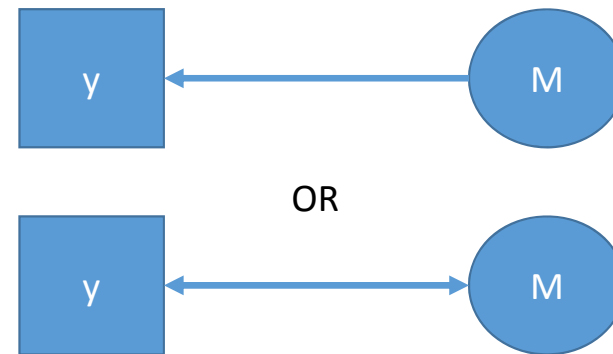
CpG Methylation in the twin model: Caveats

1. The causality is problematic



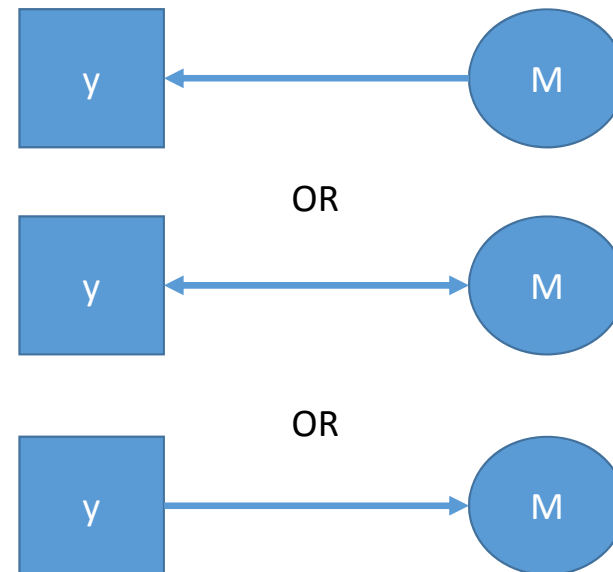
CpG Methylation in the twin model: Caveats

1. The causality is problematic



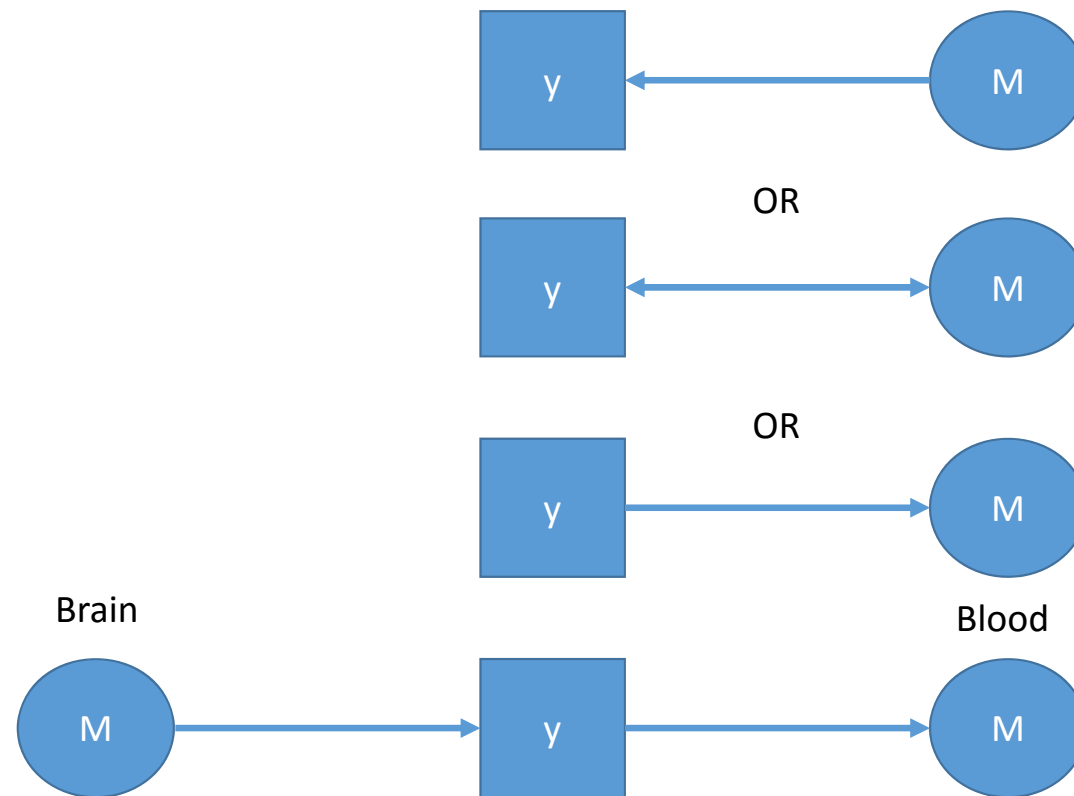
CpG Methylation in the twin model: Caveats

1. The causality is problematic



CpG Methylation in the twin model: Caveats

1. The causality is problematic

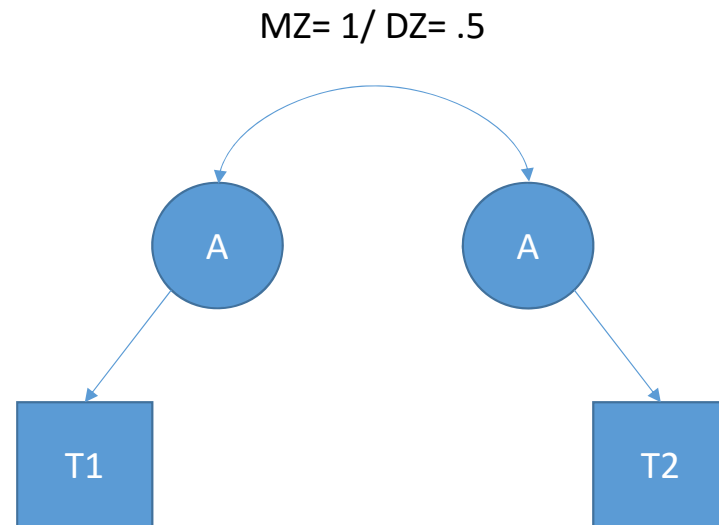


CpG Methylation in the twin model: Caveats

1. The causality is problematic
2. The DZ and MZ might behave largely additive

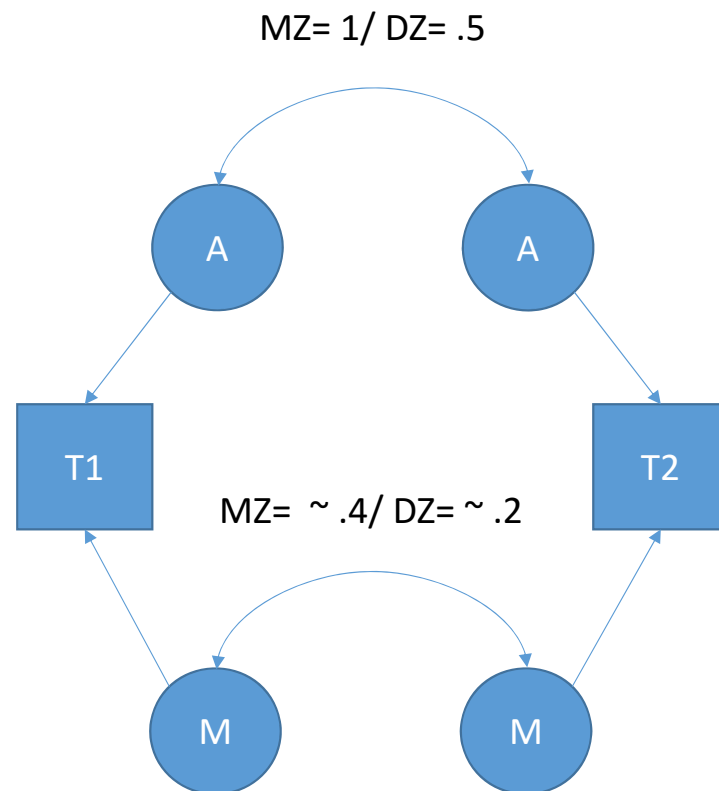
CpG Methylation in the twin model: Caveats

1. The causality is problematic
2. The DZ and MZ might behave largely additive



CpG Methylation in the twin model: Caveats

1. The causality is problematic
2. The DZ and MZ might behave largely additive
3. Methylation varies over time



CpG Methylation in the twin model: Caveats

1. The causality is problematic
2. The DZ and MZ might behave largely additive
3. Methylation varies over time

Anxiety
wave 1

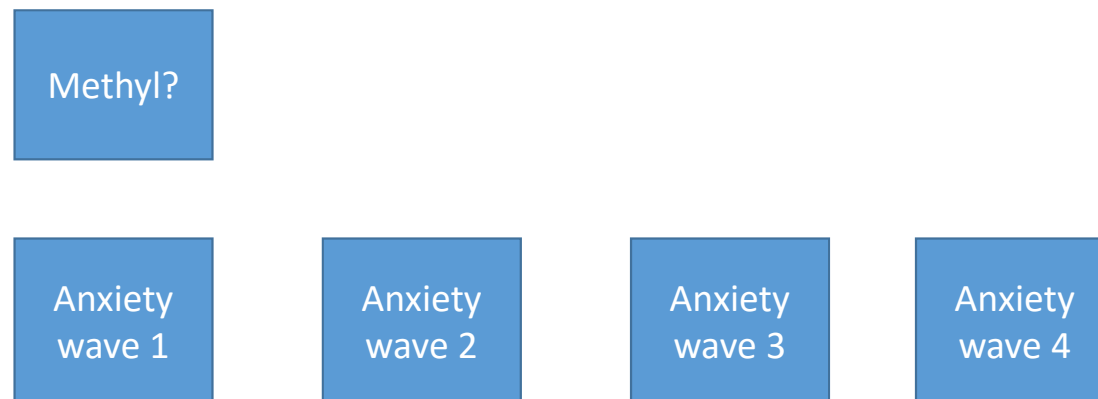
Anxiety
wave 2

Anxiety
wave 3

Anxiety
wave 4

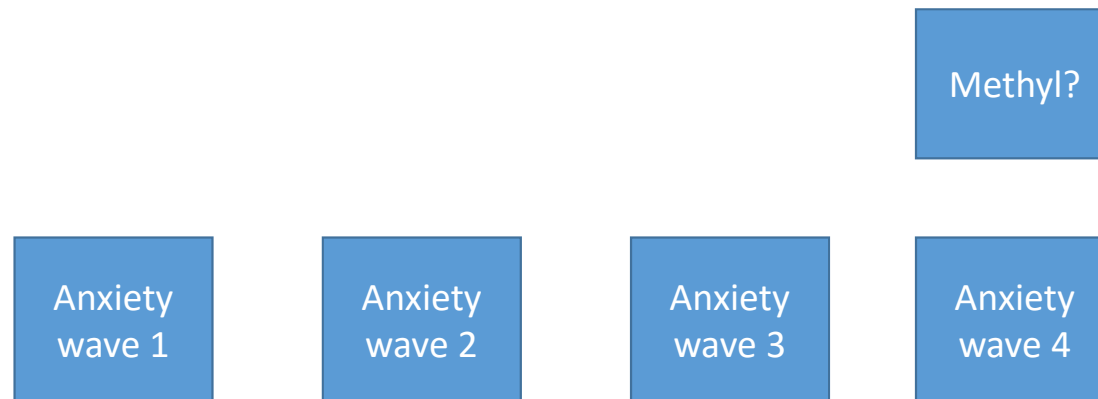
CpG Methylation in the twin model: Caveats

1. The causality is problematic
2. The DZ and MZ might behave largely additive
3. Methylation varies over time



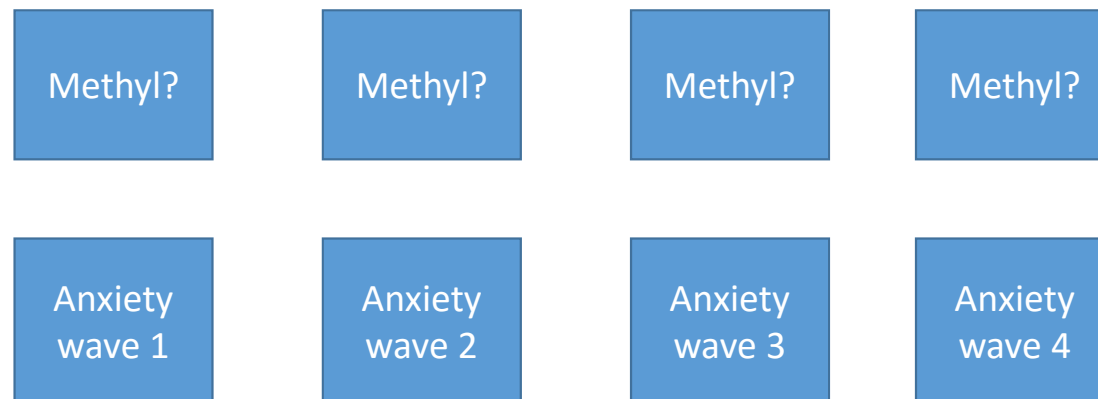
CpG Methylation in the twin model: Caveats

1. The causality is problematic
2. The DZ and MZ might behave largely additive
3. Methylation varies over time

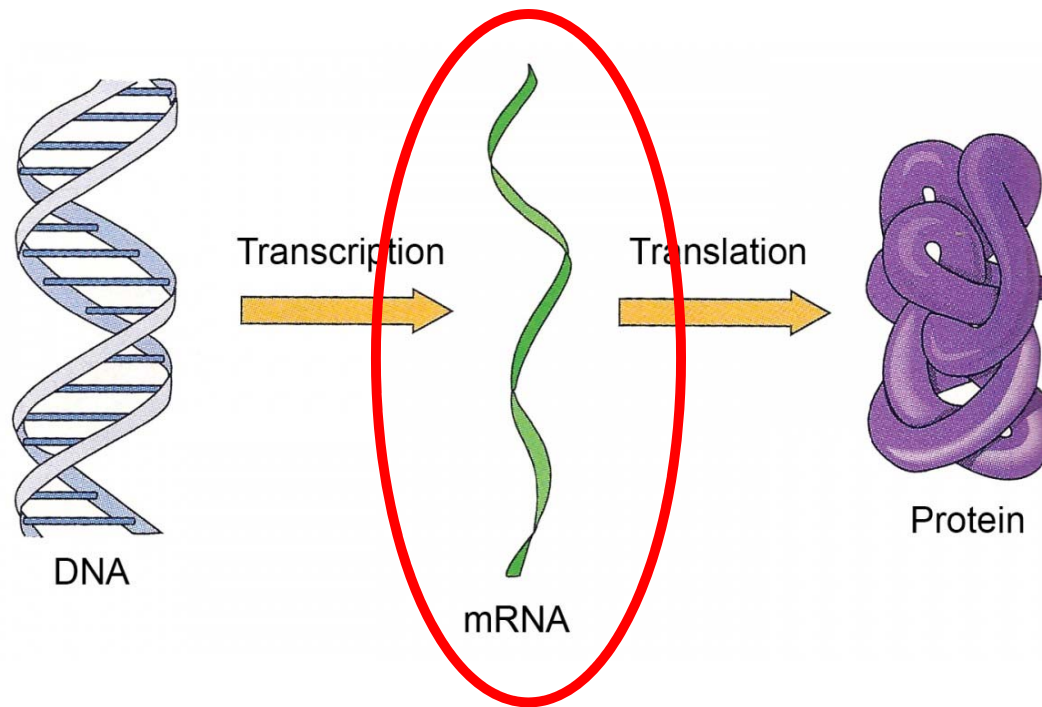


CpG Methylation in the twin model: Caveats

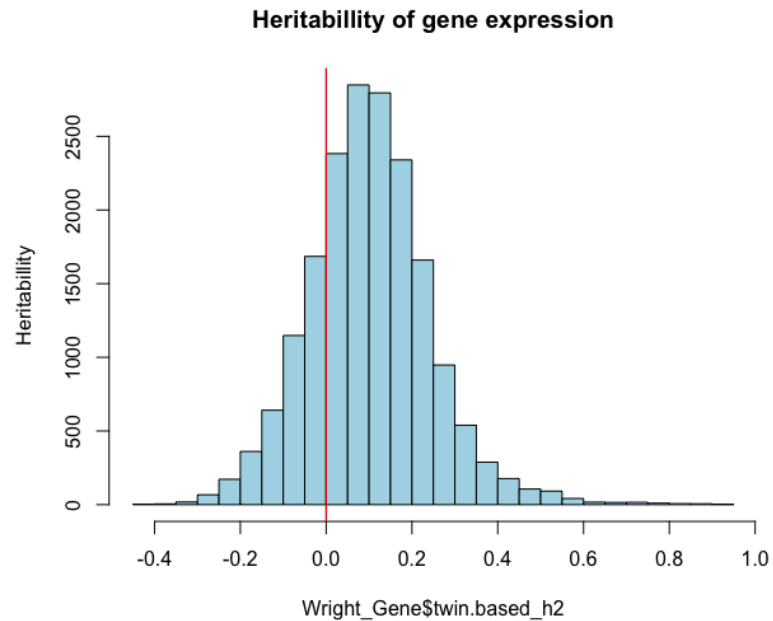
1. The causality is problematic
2. The DZ and MZ might behave largely additive
3. Methylation varies over time



The next step: Gene expression



Heritability of gene expression



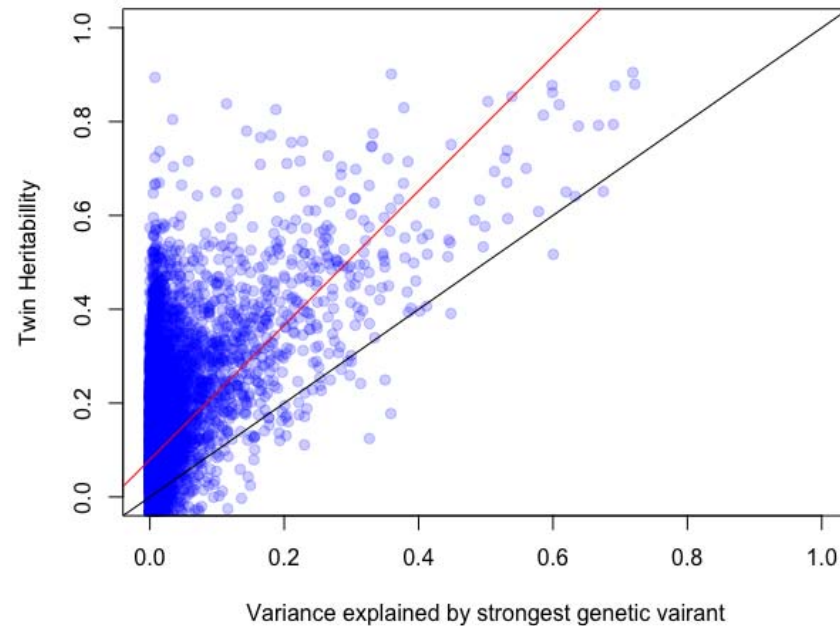
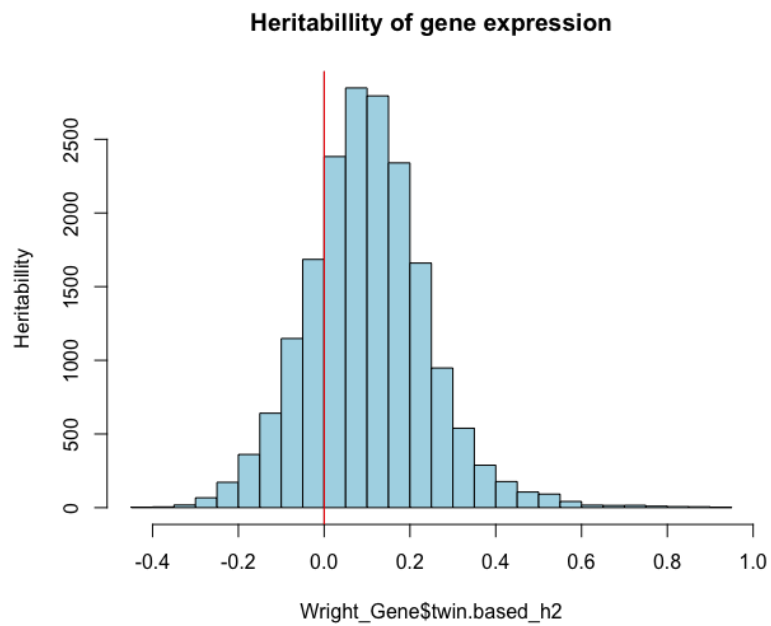
Wright et al. (2014):

Assessed gene expression profiles in 2,752 twins

Assessed the relationship between Gene expression and DNA sequence

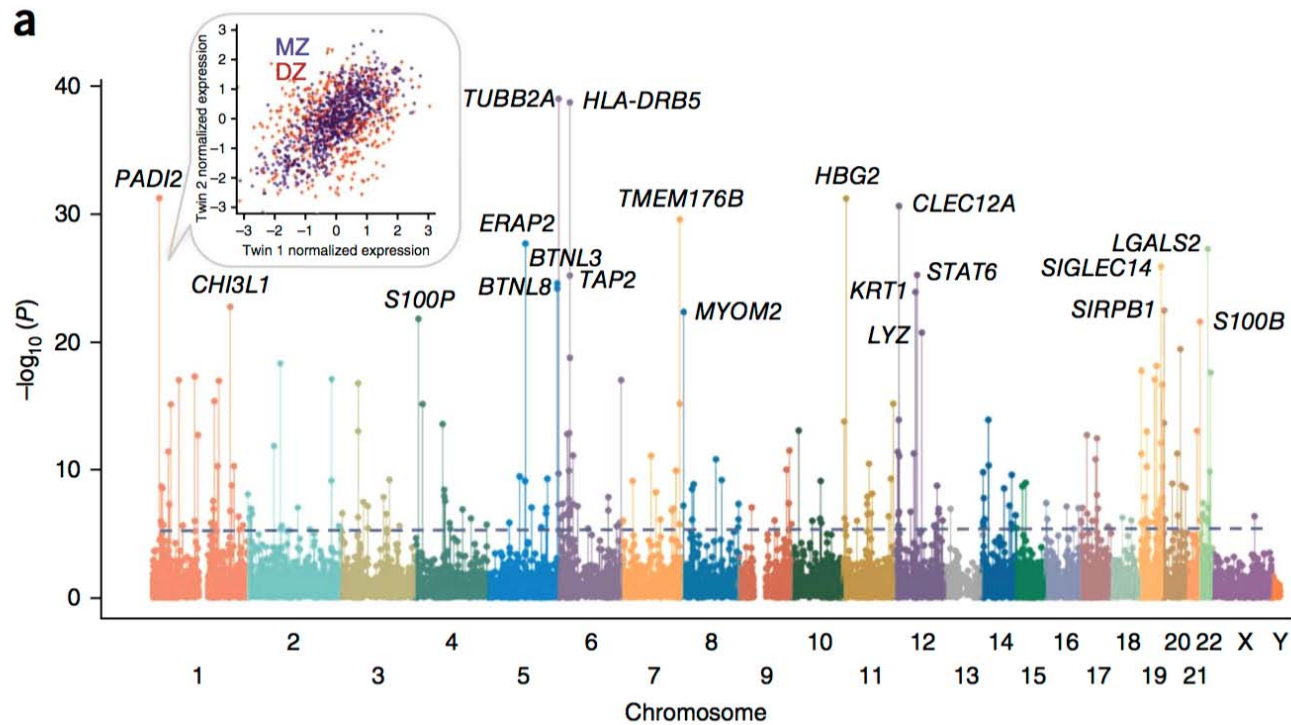
Wright, F. A., Sullivan, P. F., Brooks, A. I., Zou, F., Sun, W., Xia, K., ... & Abdellaoui, A. (2014). Heritability and genomics of gene expression in peripheral blood. *Nature genetics*, 46(5), 430-437.

Heritability of gene expression



Wright, F. A., Sullivan, P. F., Brooks, A. I., Zou, F., Sun, W., Xia, K., ... & Abdellaoui, A. (2014). Heritability and genomics of gene expression in peripheral blood. *Nature genetics*, 46(5), 430-437.

Heritability of gene expression



Wright, F. A., Sullivan, P. F., Brooks, A. I., Zou, F., Sun, W., Xia, K., ... & Abdellaoui, A. (2014). Heritability and genomics of gene expression in peripheral blood. *Nature genetics*, 46(5), 430-437.

Heritability of gene expression

Table 2 Predictors of high heritability expression levels

Predictor	Mean h^2			Expression-corrected	
	change	Enrichment z	P	enrichment z	P
Mean expression		11.25	2.43×10^{-29}	–	–
Variance in expression		14.14	2.23×10^{-45}	14.89	4.02×10^{-50}
GC content, +5 kb of TSS		–1.42	0.155	–5.33	9.60×10^{-8}
GC content, –5 kb of TSS		–0.72	0.471	–5.00	5.73×10^{-7}
DHS near TSS ^a		9.45	3.55×10^{-21}	4.01	6.00×10^{-5}
DHS near TSS, blood		8.87	7.02×10^{-16}	1.30	0.195
Gene density ^b		–6.98	2.98×10^{-12}	–10.85	2.09×10^{-27}
Gene size ^c		8.07	7.02×10^{-16}	11.30	1.27×10^{-29}
Local recombination rate ^d		0.73	0.464	3.01	0.0026
Size of LD block ^e		–0.05	0.959	–0.49	0.622
Gene conservation score ^f		8.49	2.00×10^{-17}	1.14	0.255
Genes under selection (185) ^g	0.013	1.60	0.109	1.82	0.068
Genes under positive selection (549) ^h	0.007	1.32	0.186	1.78	0.074
Genes under balancing selection (47) ⁱ	0.042	2.65	0.0081	2.83	0.0046
Genes under adaptive selection (174) ^j	0.019	2.26	0.024	1.13	0.260
Human accelerated genes (161) ^k	0.024	3.05	0.0023	4.12	3.73×10^{-5}
Primate accelerated genes (137) ^k	0.024	2.86	0.0042	3.97	7.11×10^{-5}
NHGRI GWAS catalog (2,343) ^l	0.018	7.42	1.14×10^{-13}	7.52	5.53×10^{-14}
NHGRI, chr. 6 genes removed (2,142)	0.016	6.06	1.37×10^{-9}	6.42	1.36×10^{-10}
NHGRI, immune diseases (720) ^m	0.032	7.22	5.02×10^{-13}	5.77	7.99×10^{-9}
NHGRI, non-immune diseases (1,623)	0.011	3.71	0.0002	4.88	1.03×10^{-6}
OMIM disease entries (3,089) ⁿ	0.018	8.87	7.63×10^{-19}	7.54	4.81×10^{-14}
NHGRI + OMIM (4,809)	0.019	10.81	2.96×10^{-27}	9.84	7.81×10^{-23}

Not all genes are equally heritable!

Wright, F. A. et al (2014)

Heritability of gene expression

Table 2 Predictors of high heritability expression levels

Predictor	Mean h^2			Expression-corrected	
	change	Enrichment z	P	enrichment z	P
Mean expression		11.25	2.43×10^{-29}	–	–
Variance in expression		14.14	2.23×10^{-45}	14.89	4.02×10^{-50}
GC content, +5 kb of TSS		–1.42	0.155	–5.33	9.60×10^{-8}
GC content, –5 kb of TSS		–0.72	0.471	–5.00	5.73×10^{-7}
DHS near TSS ^a		9.45	3.55×10^{-21}	4.01	6.00×10^{-5}
DHS near TSS, blood		8.87	7.02×10^{-16}	1.30	0.195
Gene density ^b		–6.98	2.98×10^{-12}	–10.85	2.09×10^{-27}
Gene size ^c		8.07	7.02×10^{-16}	11.30	1.27×10^{-29}
Local recombination rate ^d		0.73	0.464	3.01	0.0026
Size of LD block ^e		–0.05	0.959	–0.49	0.622
Gene conservation score ^f		8.49	2.00×10^{-17}	1.14	0.255
Genes under selection (185) ^g	0.013	1.60	0.109	1.82	0.068
Genes under positive selection (549) ^h	0.007	1.32	0.186	1.78	0.074
Genes under balancing selection (47) ⁱ	0.042	2.65	0.0081	2.83	0.0046
Genes under adaptive selection (174) ^j	0.019	2.26	0.024	1.13	0.260
Human accelerated genes (161) ^k	0.024	3.05	0.0023	4.12	3.73×10^{-5}
Primate accelerated genes (137) ^k	0.024	2.86	0.0042	3.97	7.11×10^{-5}
NHGRI GWAS catalog (2,343) ^l	0.018	7.42	1.14×10^{-13}	7.52	5.53×10^{-14}
NHGRI, chr. 6 genes removed (2,142)	0.016	6.06	1.37×10^{-9}	6.42	1.36×10^{-10}
NHGRI, immune diseases (720) ^m	0.032	7.22	5.02×10^{-13}	5.77	7.99×10^{-9}
NHGRI, non-immune diseases (1,623)	0.011	3.71	0.0002	4.88	1.03×10^{-6}
OMIM disease entries (3,089) ⁿ	0.018	8.87	7.63×10^{-19}	7.54	4.81×10^{-14}
NHGRI + OMIM (4,809)	0.019	10.81	2.96×10^{-27}	9.84	7.81×10^{-23}

Genes near regions that can be epigenetically Modified, are less heritable.

Heritability of gene expression

Table 2 Predictors of high heritability expression levels

Predictor	Mean h^2			Expression-corrected	
	change	Enrichment z	P	enrichment z	P
Mean expression		11.25	2.43×10^{-29}	–	–
Variance in expression		14.14	2.23×10^{-45}	14.89	4.02×10^{-50}
GC content, +5 kb of TSS		–1.42	0.155	–5.33	9.60×10^{-8}
GC content, –5 kb of TSS		–0.72	0.471	–5.00	5.73×10^{-7}
DHS near TSS ^a		9.45	3.55×10^{-21}	4.01	6.00×10^{-5}
DHS near TSS, blood		8.87	7.02×10^{-16}	1.30	0.195
Gene density ^b		–6.98	2.98×10^{-12}	–10.85	2.09×10^{-27}
Gene size ^c		8.07	7.02×10^{-16}	11.30	1.27×10^{-29}
Local recombination rate ^d		0.73	0.464	3.01	0.0026
Size of LD block ^e		–0.05	0.959	–0.49	0.622
Gene conservation score ^f		8.49	2.00×10^{-17}	1.14	0.255
Genes under selection (185) ^g	0.013	1.60	0.109	1.82	0.068
Genes under positive selection (549) ^h	0.007	1.32	0.186	1.78	0.074
Genes under balancing selection (47) ⁱ	0.042	2.65	0.0081	2.83	0.0046
Genes under adaptive selection (174) ^j	0.019	2.26	0.024	1.13	0.260
Human accelerated genes (161) ^k	0.024	3.05	0.0023	4.12	3.73×10^{-5}
Primate accelerated genes (137) ^k	0.024	2.86	0.0042	3.97	7.11×10^{-5}
NHGRI GWAS catalog (2,343) ^l	0.018	7.42	1.14×10^{-13}	7.52	5.53×10^{-14}
NHGRI, chr. 6 genes removed (2,142)	0.016	6.06	1.37×10^{-9}	6.42	1.36×10^{-10}
NHGRI, immune diseases (720) ^m	0.032	7.22	5.02×10^{-13}	5.77	7.99×10^{-9}
NHGRI, non-immune diseases (1,623)	0.011	3.71	0.0002	4.88	1.03×10^{-6}
OMIM disease entries (3,089) ⁿ	0.018	8.87	7.63×10^{-19}	7.54	4.81×10^{-14}
NHGRI + OMIM (4,809)	0.019	10.81	2.96×10^{-27}	9.84	7.81×10^{-23}

Large genes are more heritable, Genes in gene dense regions are less heritable.

Heritability of gene expression

Table 2 Predictors of high heritability expression levels

Predictor	Mean h^2			Expression-corrected	
	change	Enrichment z	P	enrichment z	P
Mean expression		11.25	2.43×10^{-29}	–	–
Variance in expression		14.14	2.23×10^{-45}	14.89	4.02×10^{-50}
GC content, +5 kb of TSS		–1.42	0.155	–5.33	9.60×10^{-8}
GC content, –5 kb of TSS		–0.72	0.471	–5.00	5.73×10^{-7}
DHS near TSS ^a		9.45	3.55×10^{-21}	4.01	6.00×10^{-5}
DHS near TSS, blood		8.87	7.02×10^{-16}	1.30	0.195
Gene density ^b		–6.98	2.98×10^{-12}	–10.85	2.09×10^{-27}
Gene size ^c		8.07	7.02×10^{-16}	11.30	1.27×10^{-29}
Local recombination rate ^d		0.73	0.464	3.01	0.0026
Size of LD block ^e		–0.05	0.959	–0.49	0.622
Gene conservation score ^f		8.49	2.00×10^{-17}	1.14	0.255
Genes under selection (185) ^g	0.013	1.60	0.109	1.82	0.068
Genes under positive selection (549) ^h	0.007	1.32	0.186	1.78	0.074
Genes under balancing selection (47) ⁱ	0.042	2.65	0.0081	2.83	0.0046
Genes under adaptive selection (174) ^j	0.019	2.26	0.024	1.13	0.260
Human accelerated genes (161) ^k	0.024	3.05	0.0023	4.12	3.73×10^{-5}
Primate accelerated genes (137) ^k	0.024	2.86	0.0042	3.97	7.11×10^{-5}
NHGRI GWAS catalog (2,343) ^l	0.018	7.42	1.14×10^{-13}	7.52	5.53×10^{-14}
NHGRI, chr. 6 genes removed (2,142)	0.016	6.06	1.37×10^{-9}	6.42	1.36×10^{-10}
NHGRI, immune diseases (720) ^m	0.032	7.22	5.02×10^{-13}	5.77	7.99×10^{-9}
NHGRI, non-immune diseases (1,623)	0.011	3.71	0.0002	4.88	1.03×10^{-6}
OMIM disease entries (3,089) ⁿ	0.018	8.87	7.63×10^{-19}	7.54	4.81×10^{-14}
NHGRI + OMIM (4,809)	0.019	10.81	2.96×10^{-27}	9.84	7.81×10^{-23}

Genes associate to disease, and near genome wide
Signals for complex traits show increased heritability

Wright, F. A. et al (2014)

How can all of this help your research?

- Tissue Gene expression:
 - <http://www.gtexportal.org/home/>
- Developmental transcriptome:
 - <http://www.brainspan.org/rnaseq/search/index.html>
- Developmental epigenome:
 - <http://www.brainspan.org/static/download.html>

Genetic epidemiology in the genomic age: Missing Heritability & The Role of Twin Studies in the Genomic Era

- The twin study has continued use in the (gen)omic era.
- Missing heritability seems not to be a massive problem.