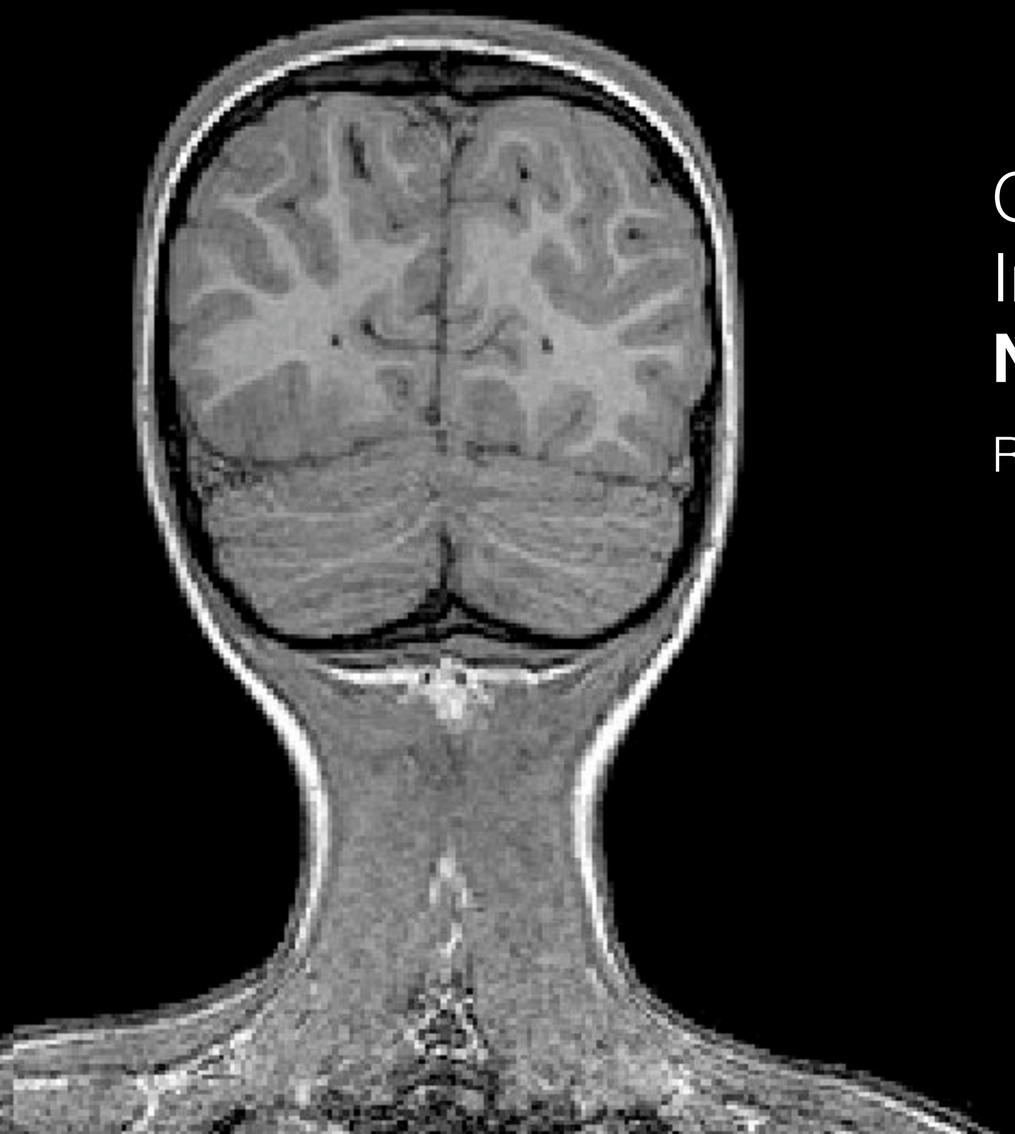
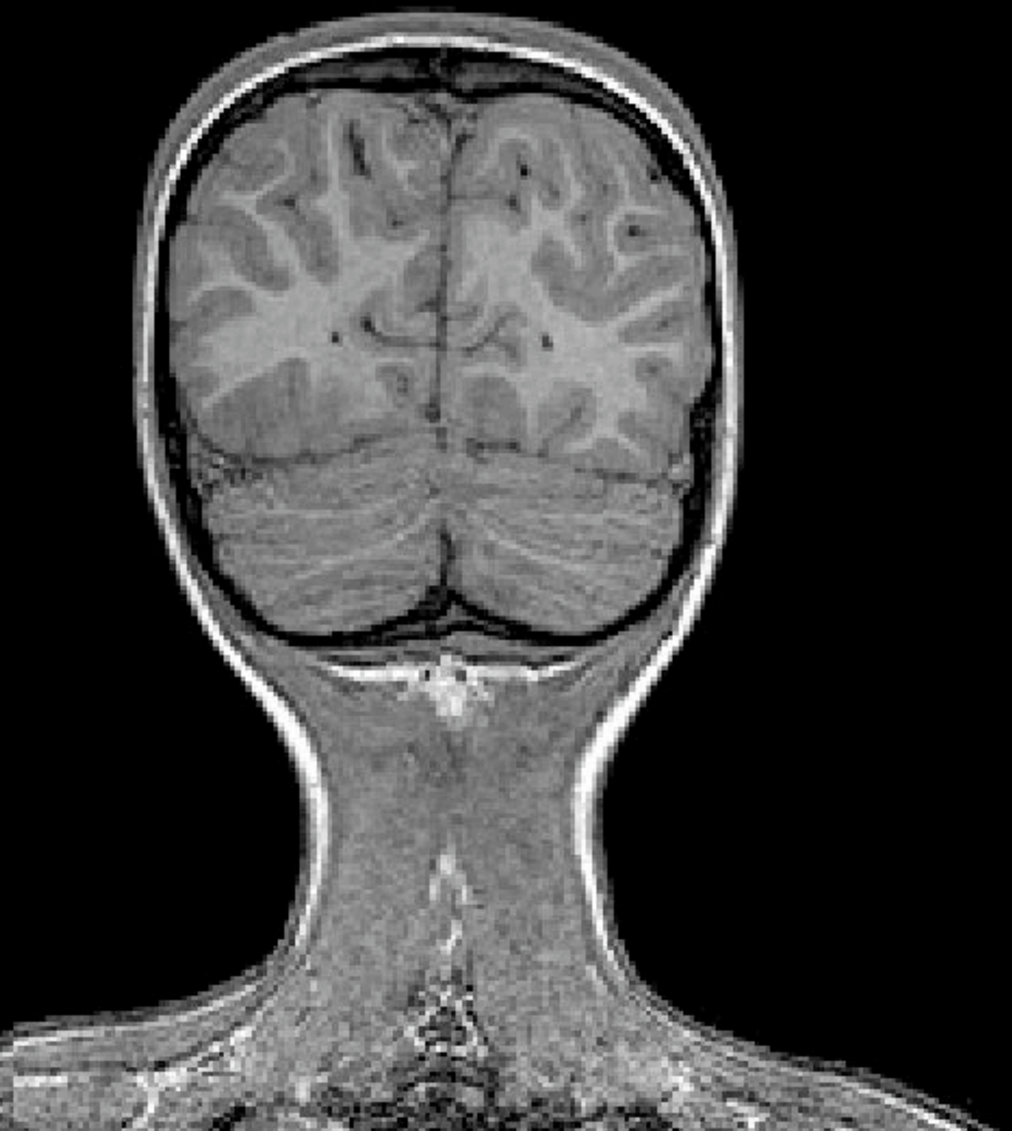


OHBM 2015

Imaging Genetics Course

Neuroimaging phenotypes and heritability

ROBERTO TORO • INSTITUT PASTEUR • rto@pasteur.fr



0.
Preliminary considerations

Phenotype and Genotype

A ***phenotype*** is any observable characteristics of an individual (biochemical, physiological, morphological,etc.) resulting from the interaction of the environment and its ***genotype*** (the genetic constitution of the individual)

Endophenotypes

In psychiatry, an endophenotype is an ***internal*** characteristic that can be ***objectively measured***, ideally in a robust and reliable fashion. An endophenotype should be ***closer to the causative biological process*** than an external phenotype (Flint&Munafò, 2007)

Glahn et al defined the ***endophenotype ranking value*** (ERV) based on the heritability of the disorder (h_i^2), the heritability (h_e^2) of the endophenotype, and their genetic correlation (ρ_g):

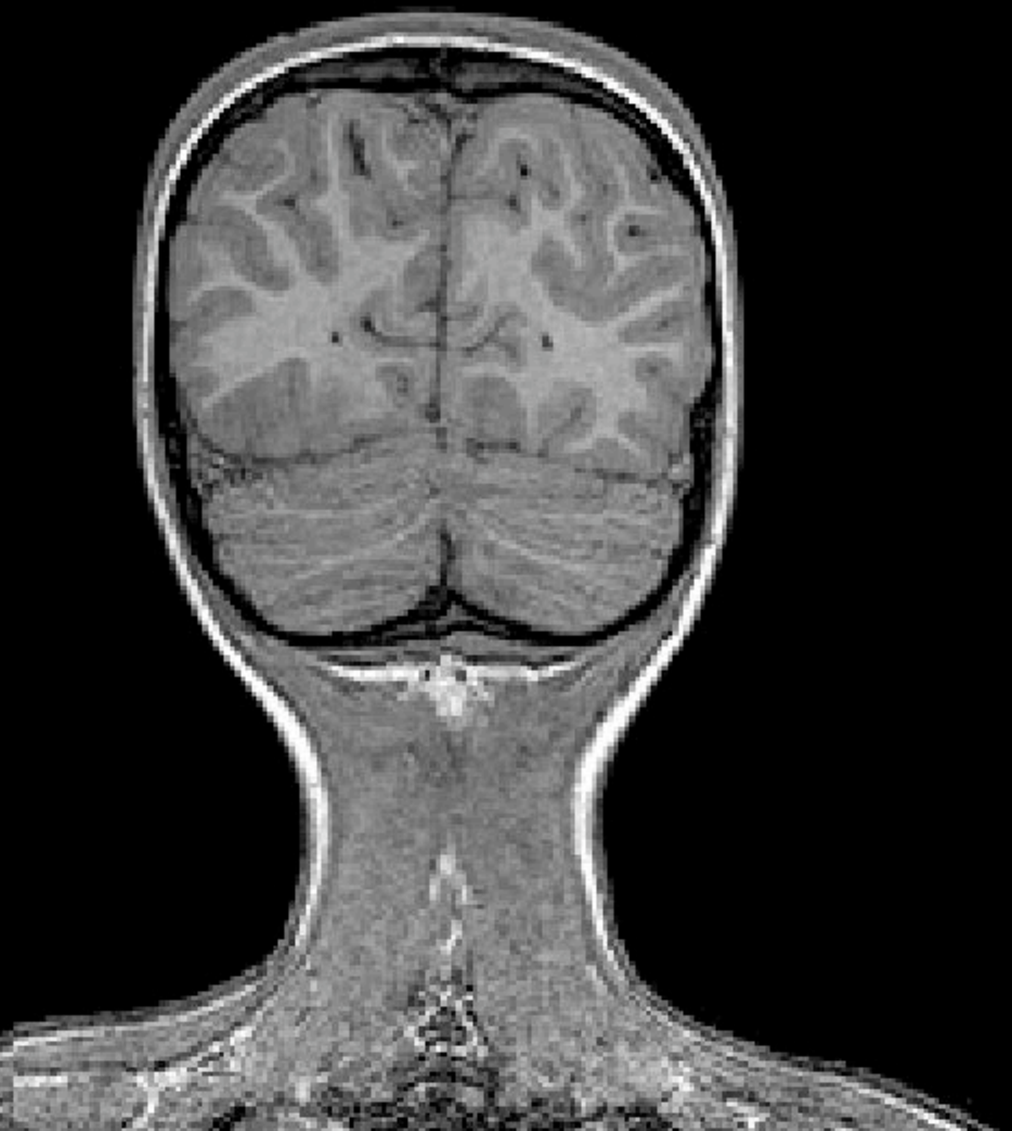
$$ERV=|\sqrt{h_i^2}\ \sqrt{h_e^2}\ \rho_g|$$

Further references

Flint and Munafò (2007) *The endophenotype concept in psychiatric genetics, Psychological Medicine, doi:10.1017/S0033291706008750*

Glahn et al (2011) *High Dimensional Endophenotype Ranking in the Search for Major Depression Risk Genes, doi:10.1016/j.biopsych.2011.08.022*

Meyer-Lindenberg and Weinberger (2006) *Intermediate phenotypes and genetic mechanisms of psychiatric disorders, doi:10.1038/nrn1993*



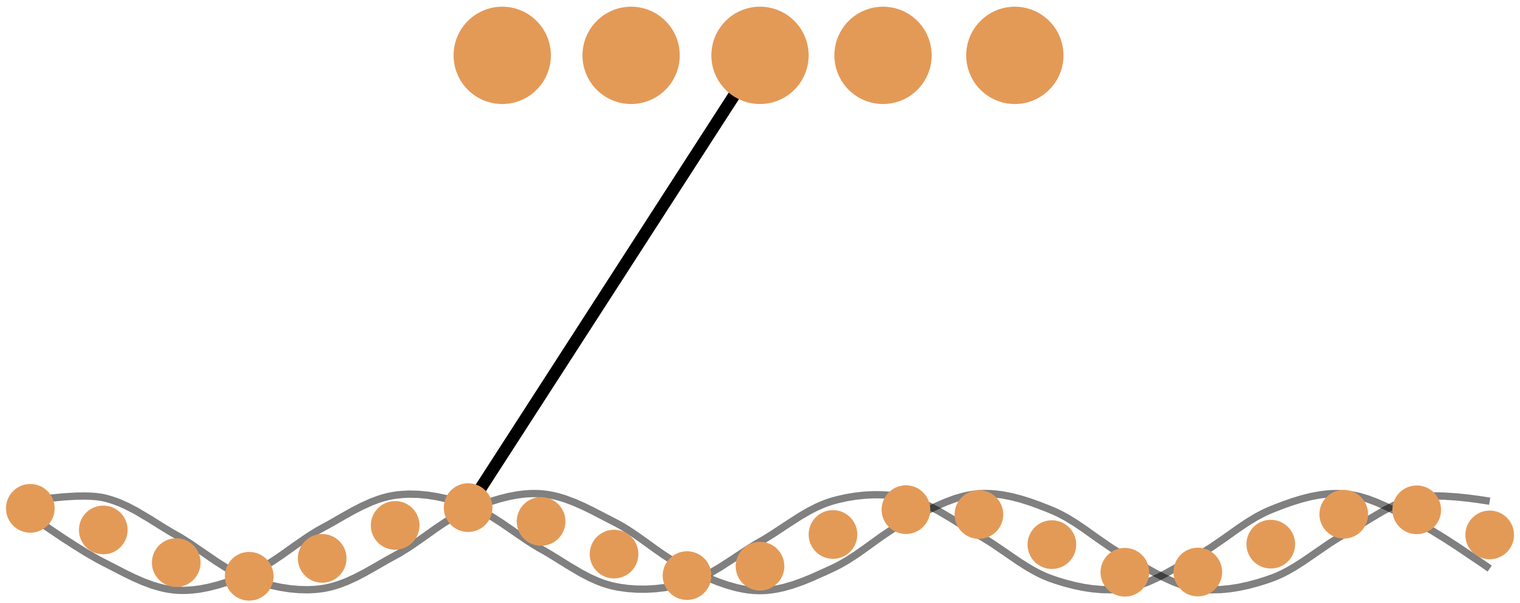
1.
Inheritance and heritability

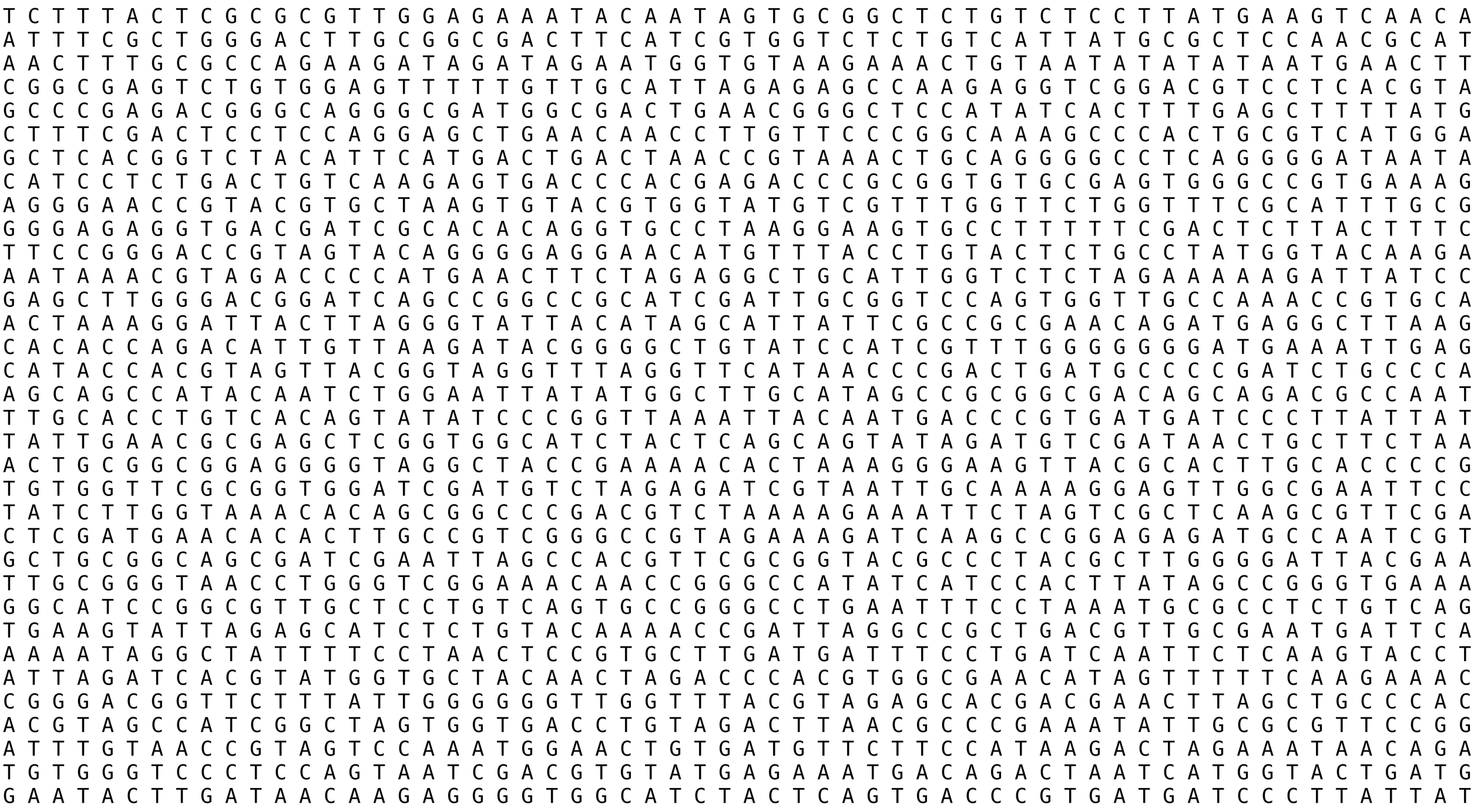
Genetic architecture of a phenotype

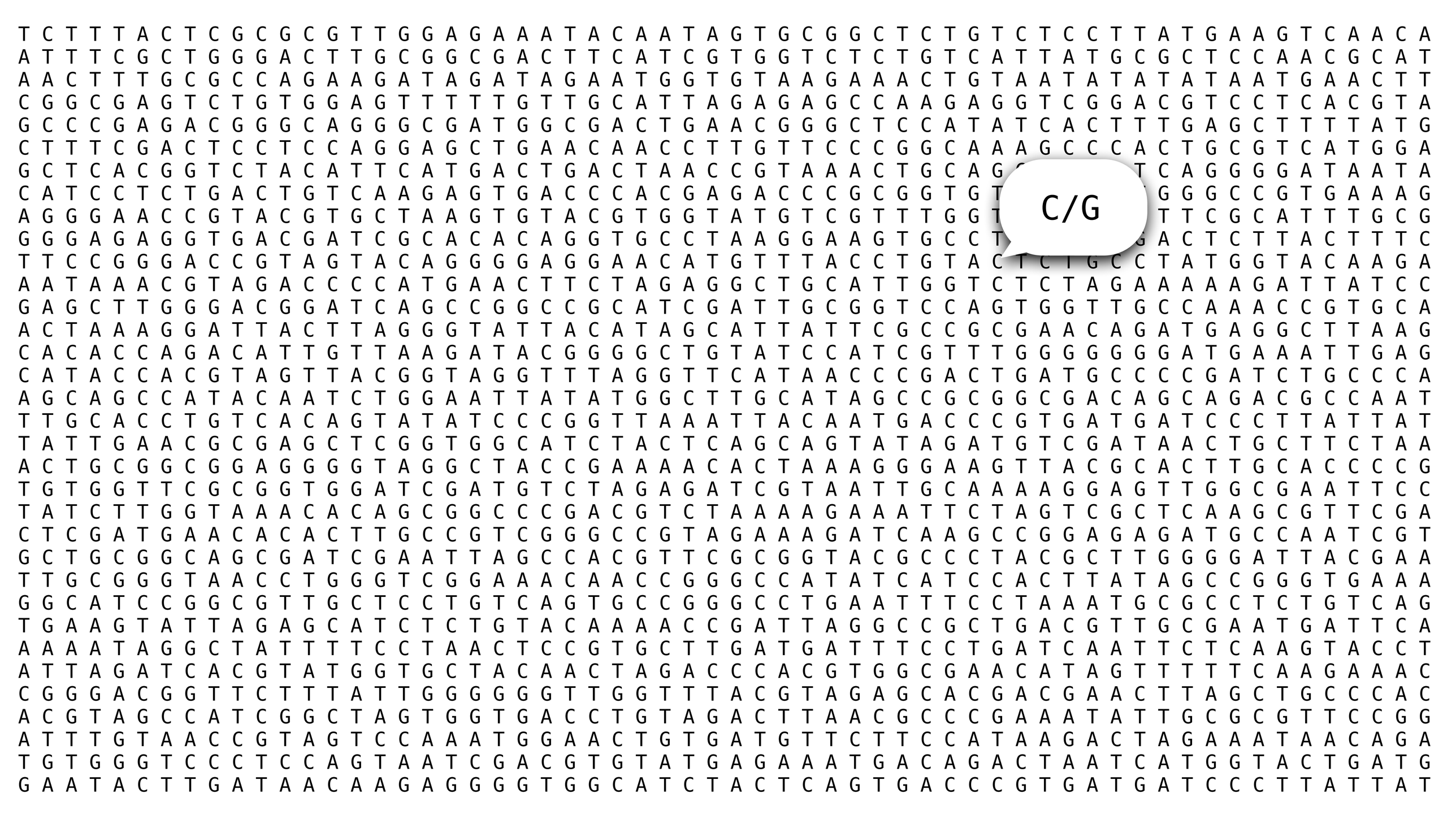
Gregor Mendel
&
Ronald Fischer

Genetic architecture of a phenotype

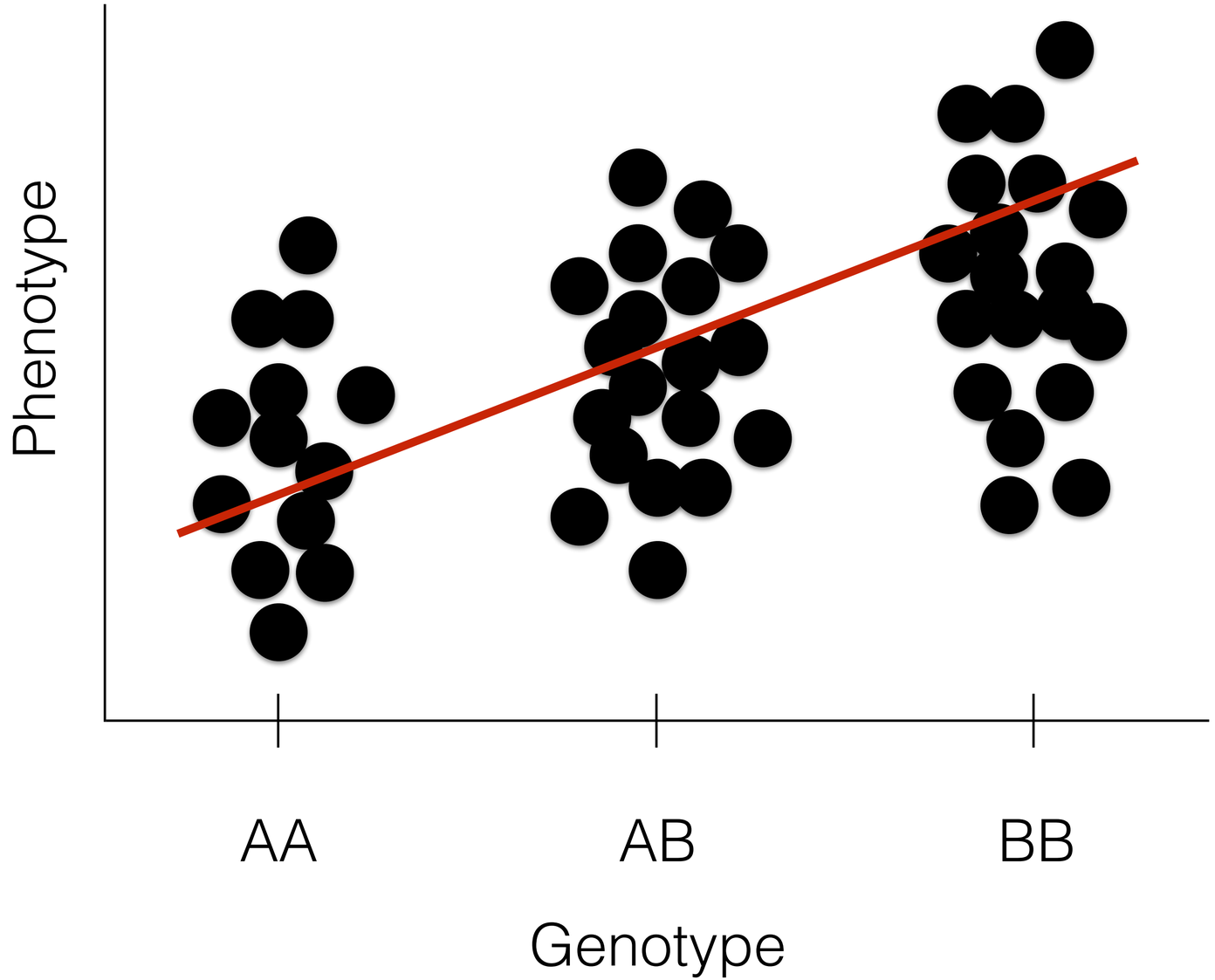
Mendelian





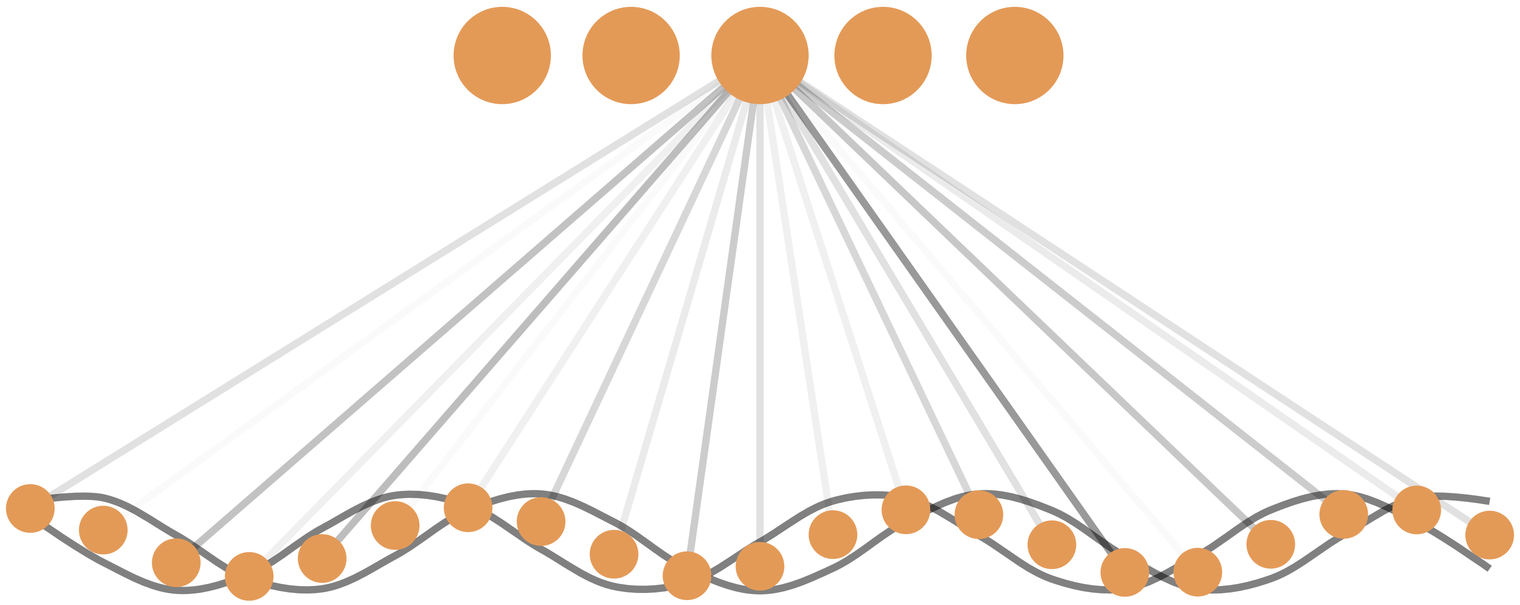

C/G

Genetic architecture of a phenotype



Genetic architecture of a phenotype

Infinitesimal
(Fischer)

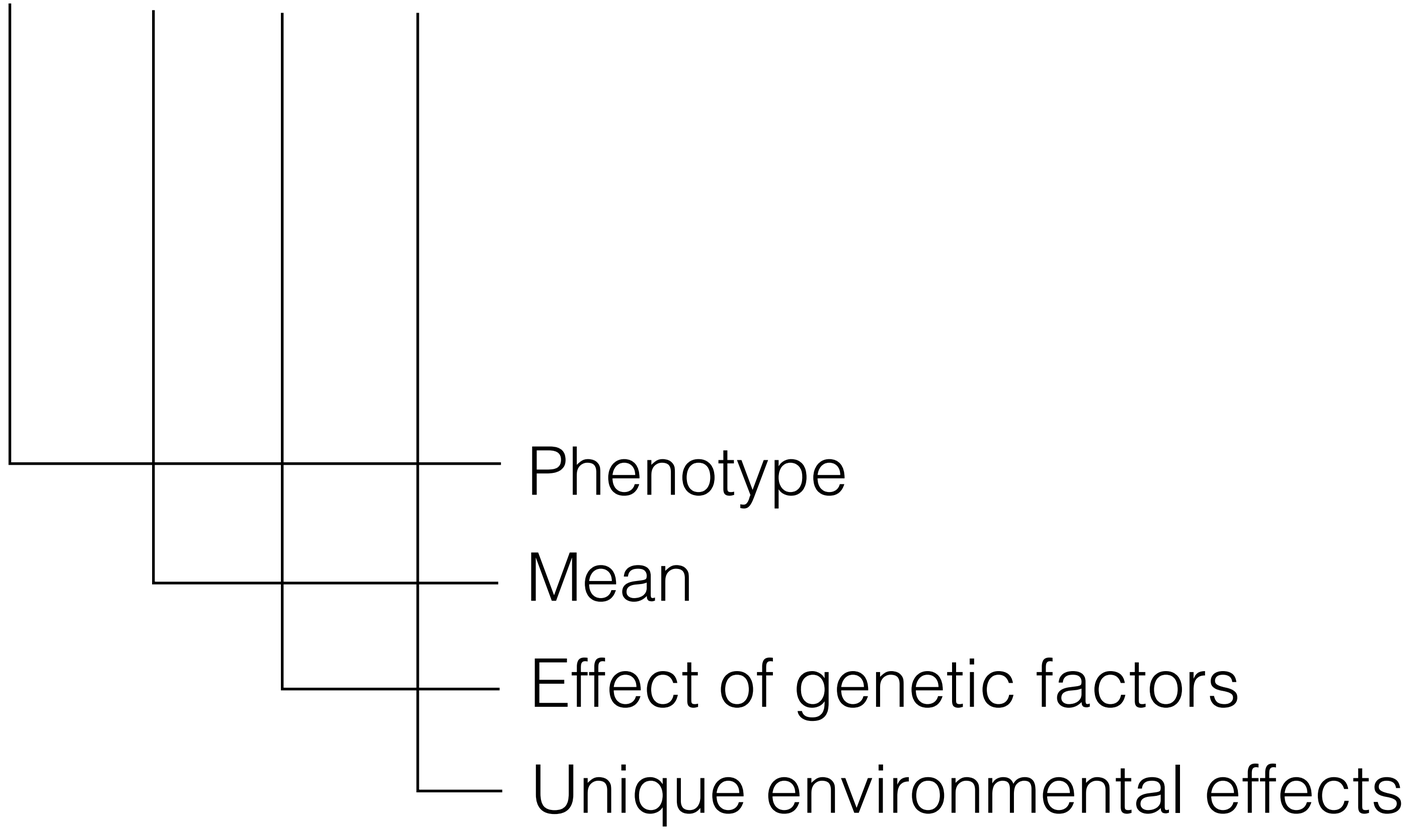


How do we study an infinitesimal genetic architecture?

Genetic architecture of a phenotype

If $\mu = E(P)$ and $G = E(P|genotype) - \mu$

then $P = \mu + G + E$



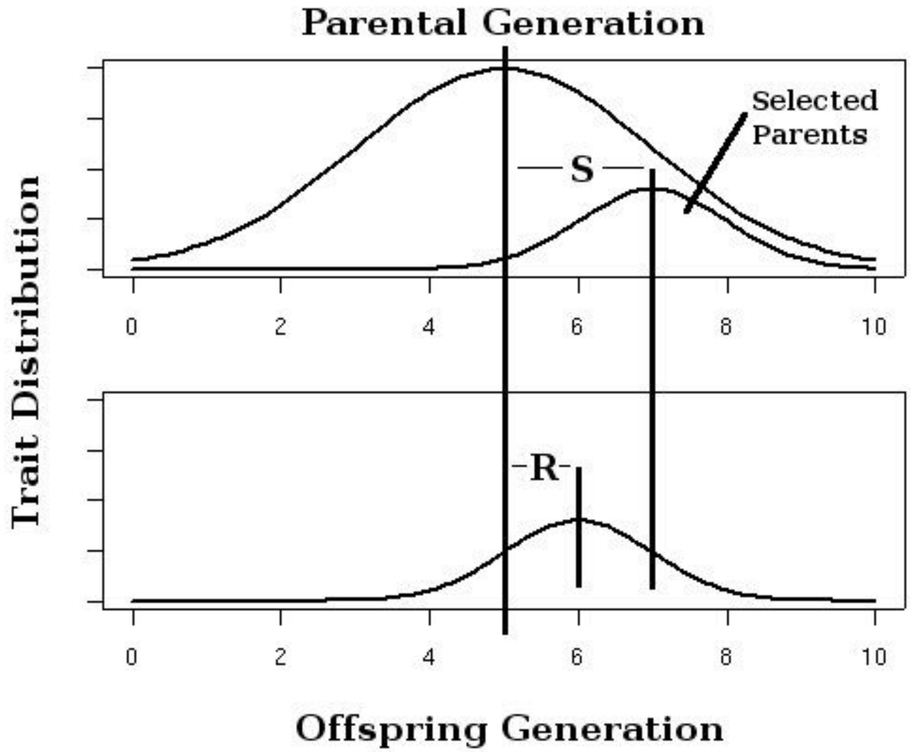
Genetic architecture of a phenotype

$$Var(P) = Var(G) + Var(E)$$

$$h^2 = \frac{Var(G)}{Var(P)} = \frac{Var(G)}{Var(G) + Var(E)}$$

The estimation of heritability allows us to study the role of genetic factors without having direct access to them

Genetic architecture of a phenotype

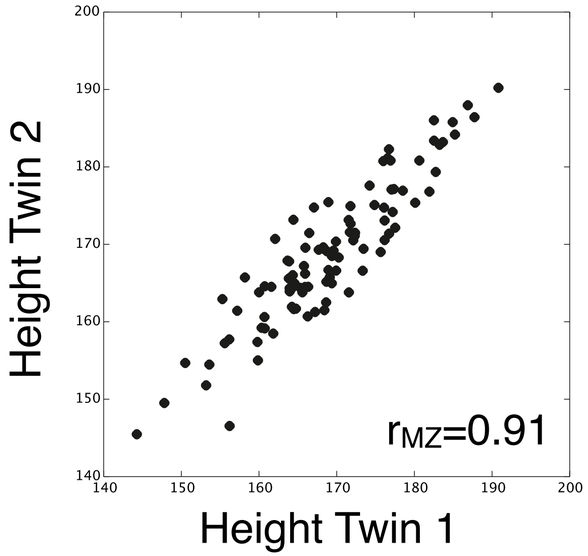


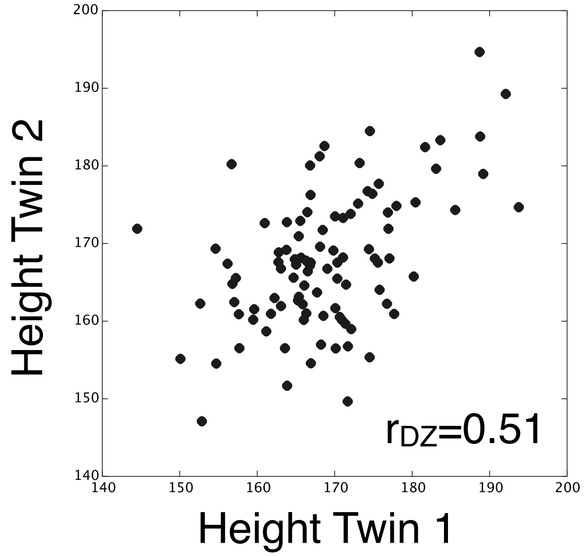
$$h^2 = \frac{R}{S}$$

Response to selection

Selection differential

Genetic architecture of a phenotype





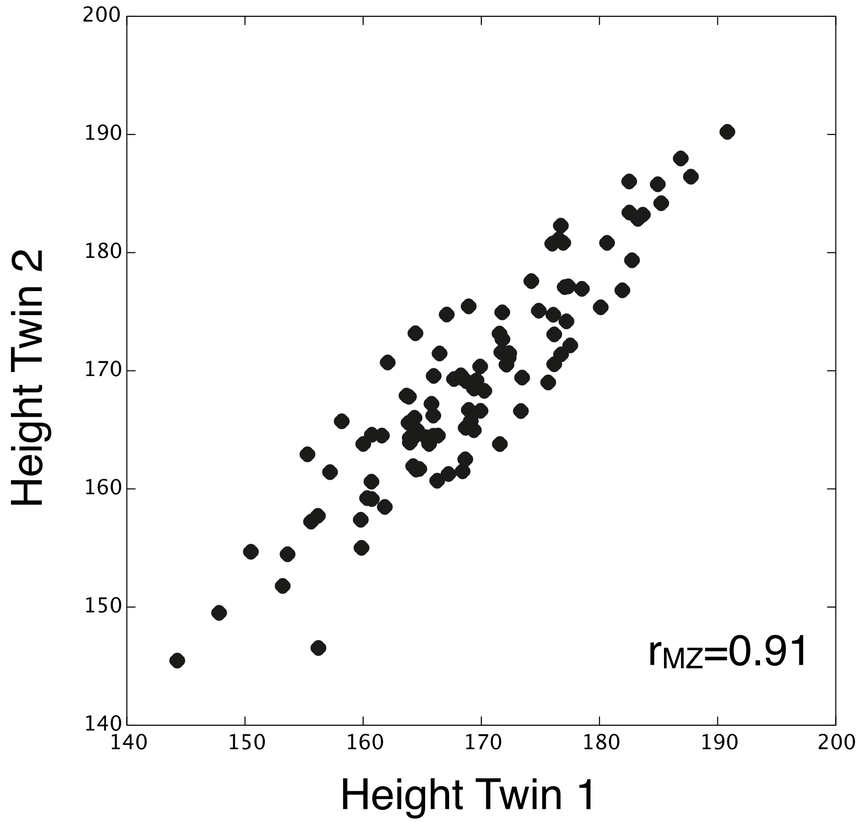
$r =$ Inheritable factors + Shared environment

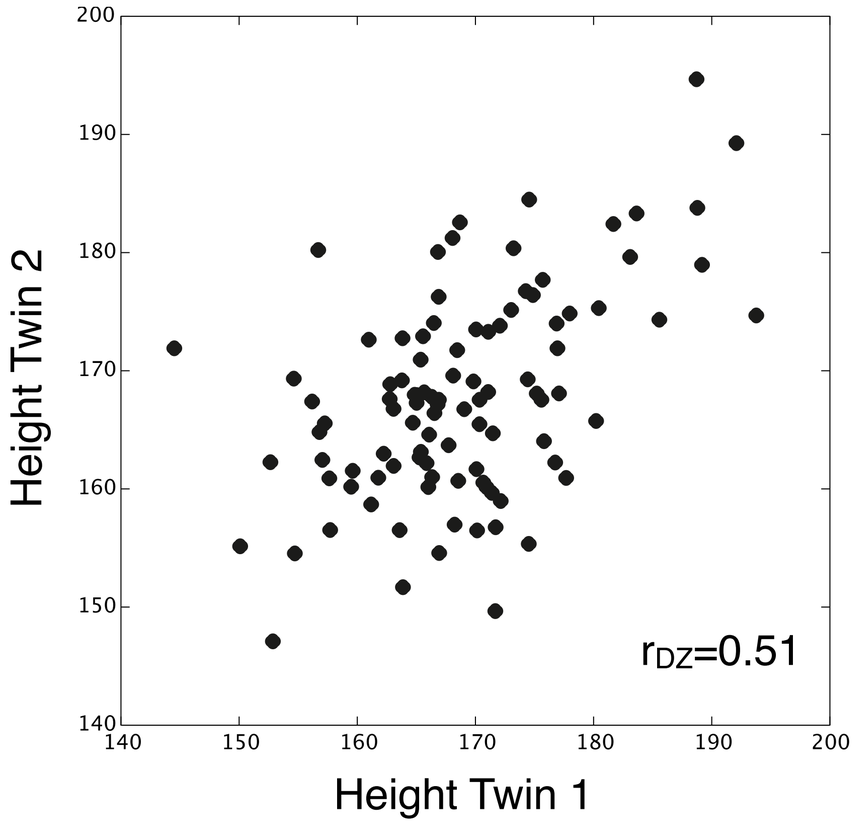
$$r_{MZ} = A + C$$

$$r_{DZ} = \frac{1}{2}A + C$$

$$A = h^2 = 2(r_{MZ} - r_{DZ})$$
$$C = r_{MZ} - A$$
$$E = 1 - A$$

Genetic architecture of a phenotype





$$h^2 = 2(0.91 - 0.51) = 0.8$$

Genetic architecture of a phenotype

$$r_{MZ} = (1)A + C$$
$$r_{DZ} = (1/2)A + C$$

$$r_{ij} = \alpha A + C$$

the idea behind Genomic Complex Trait Analysis – GCTA

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]

SNPs discovered by genome-wide association studies (GWASs) account for only a small fraction of the genetic variation of complex traits in human populations. Where is the remaining heritability? We estimated the proportion of variance for human height explained by 294,831 SNPs genotyped on 3,925 unrelated individuals using a linear model analysis, and validated the estimation method with simulations based on the observed genotype data. We show that 45% of variance can be explained by considering all SNPs simultaneously. Thus, most of the heritability is not missing but has not previously been detected because the individual effects are too small to pass stringent significance tests. We provide evidence that the remaining heritability is due to incomplete linkage disequilibrium between causal variants and genotyped SNPs, exacerbated by causal variants having lower minor allele frequency than the SNPs explored to date.

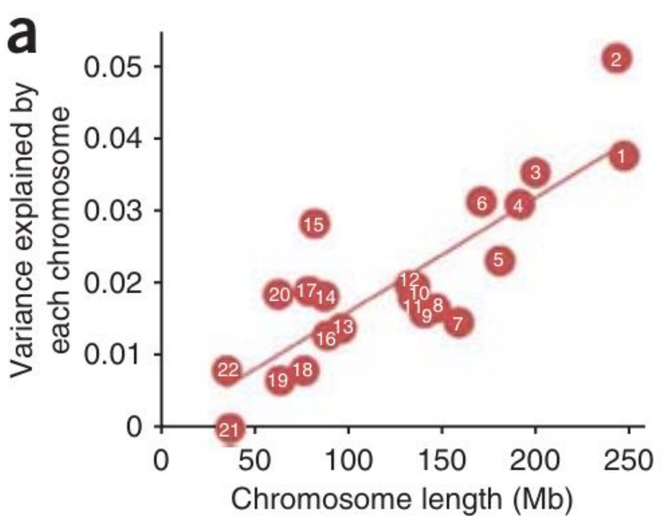
of variation that their effects do not reach stringent significance thresholds and/or the causal variants are not in complete linkage disequilibrium (LD) with the SNPs that have been genotyped. Lack of complete LD might, for instance, occur if causal variants have lower minor allele frequency (MAF) than genotyped SNPs. Here we test these two hypotheses and estimate the contribution of each to the heritability of height in humans as a model complex trait.

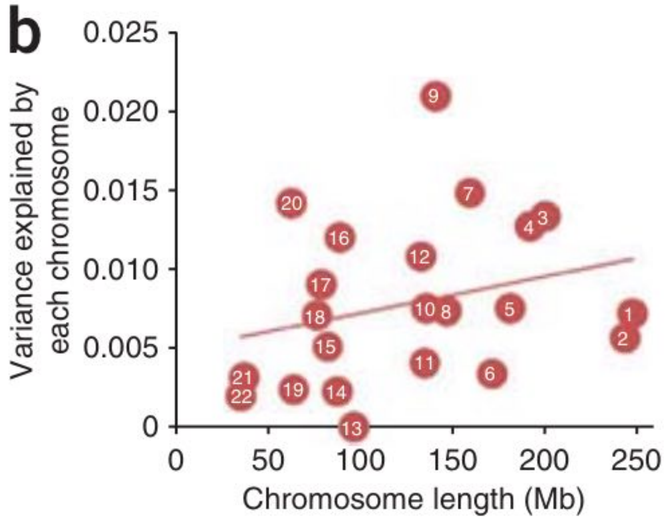
Height in humans is a classical quantitative trait, easy to measure and studied for well over a century as a model for investigating the genetic basis of complex traits[9,10]. The heritability of height has been estimated to be ~0.8 (refs. 9,11–13). Rare mutations that cause extreme short or tall stature have been found[14,15], but these do not explain much of the variation in the general population. Recent GWASs on tens of thousands of individuals have detected ~50 variants that are associated with height in the population, but these in total account for only ~5% of phenotypic variance[16–19].

Data from a GWAS that are collected to detect statistical associations

The genomic architecture of many quantitative traits is strongly polygenic

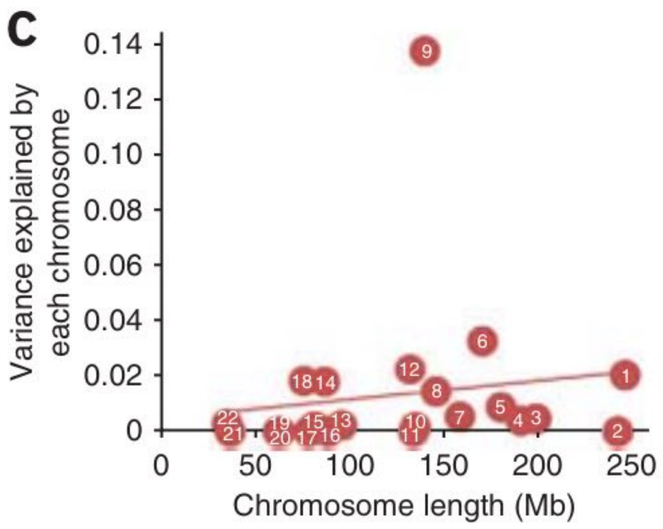
Height

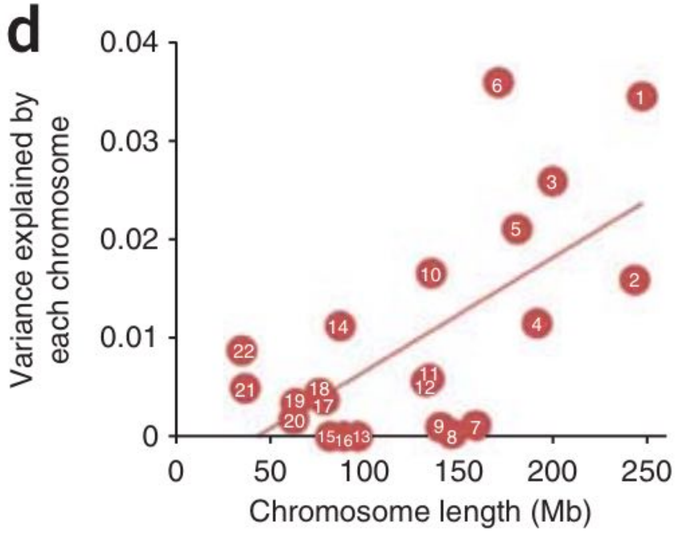




BMI

vWF





QTi

Yang et al 2011, 2013

Genome-wide SNPs capture:

45% of Height diversity (Yang et al, Nat Genet 2010)

51% of IQ diversity (Davies et al, Mol Psychiatry, 2011)

50% of neuroanatomical diversity (Toro et al, Mol Psychiatry, 2014)

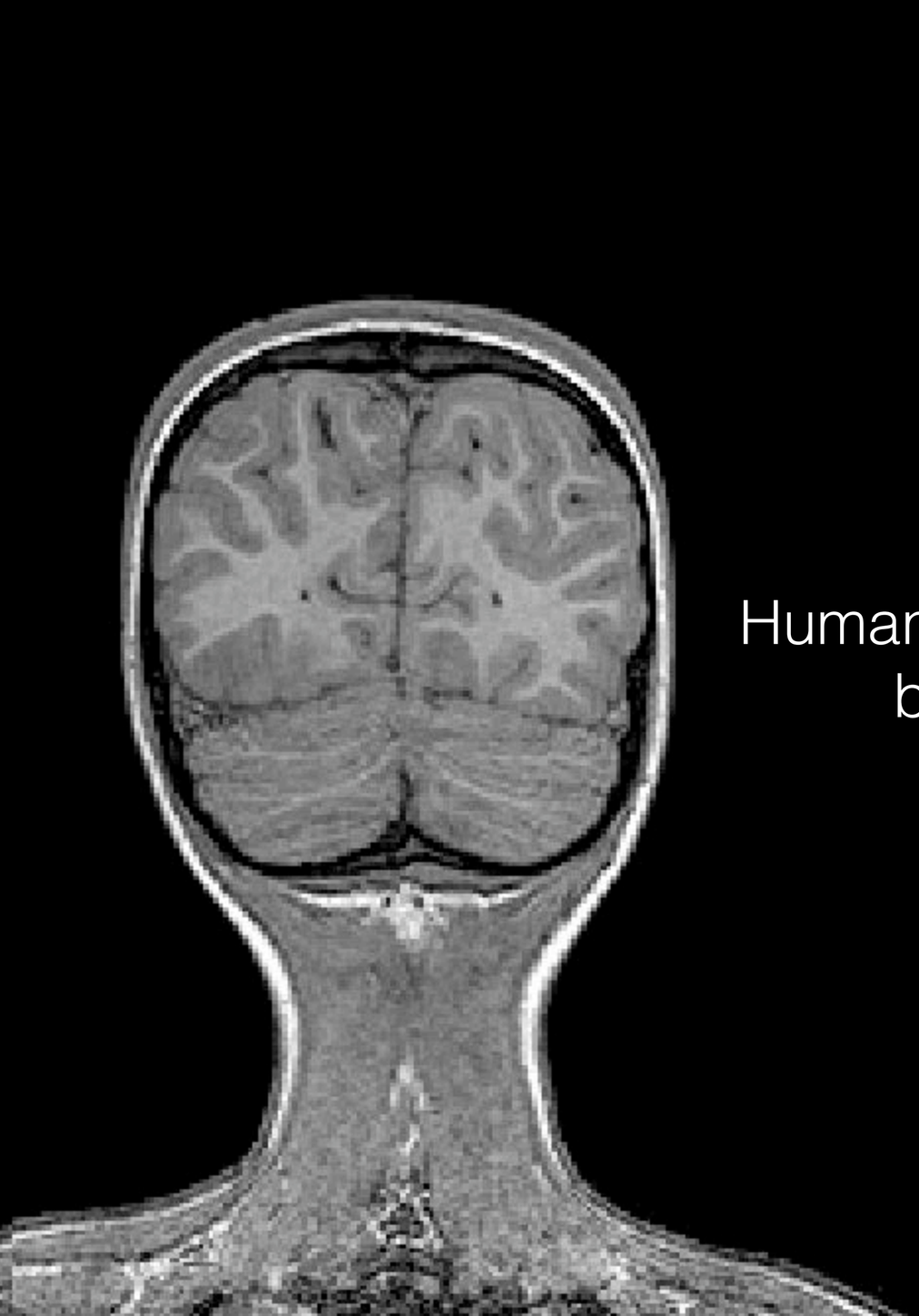
from **8% to 77%** of 49 disease-relevant traits (Yang et al, PLoS Genet 2013)

But also

23% of the risk to Schizophrenia (Lee et al, Nat Genet 2012).

24% of the risk to Alzheimer's disease (Lee et al, Hum Mol Genet 2013),

50% of the risk to Autism spectrum disorders (Gaugler et al, Nat Genet 2014)



Human brain diversity and pathology may be encoded by **tens of thousands** of genomic variants.

How to get enough statistical power?

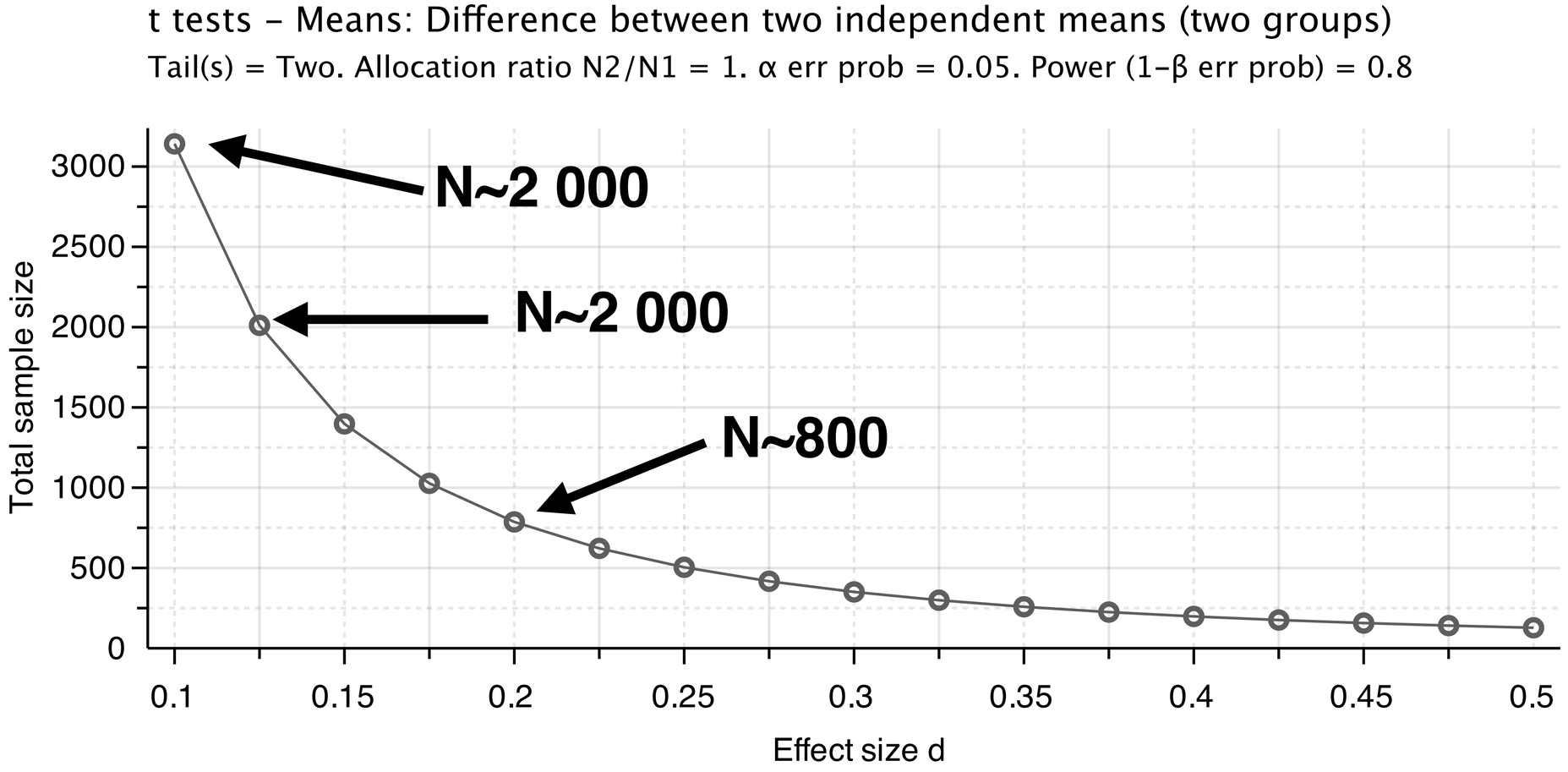
Statistical Power

Mendelian hypothesis, candidate gene

Statistical Power

Mendelian hypothesis, candidate gene

→G*Power (http://www.gpower.hhu.de)



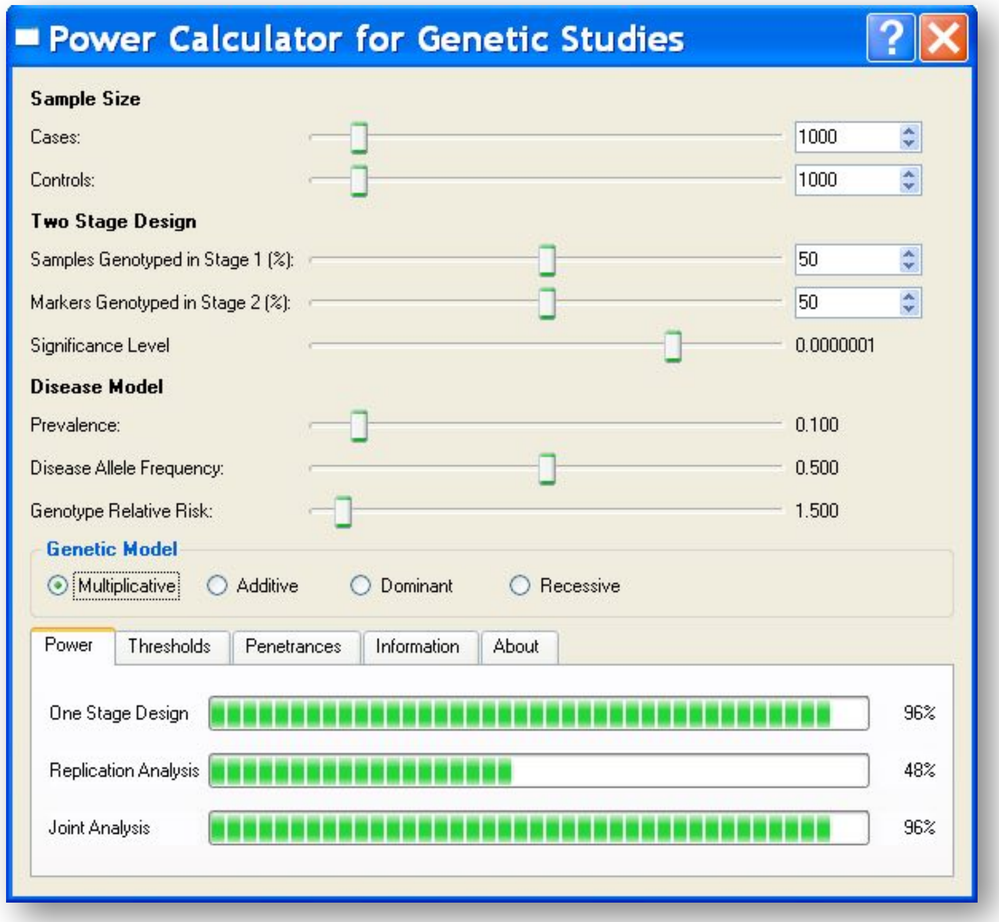
Statistical Power

Mendelian hypothesis, GWAS

Statistical Power

Mendelian hypothesis, GWAS

→ CaTS (http://www.sph.umich.edu/csg/abecasis/cats/)



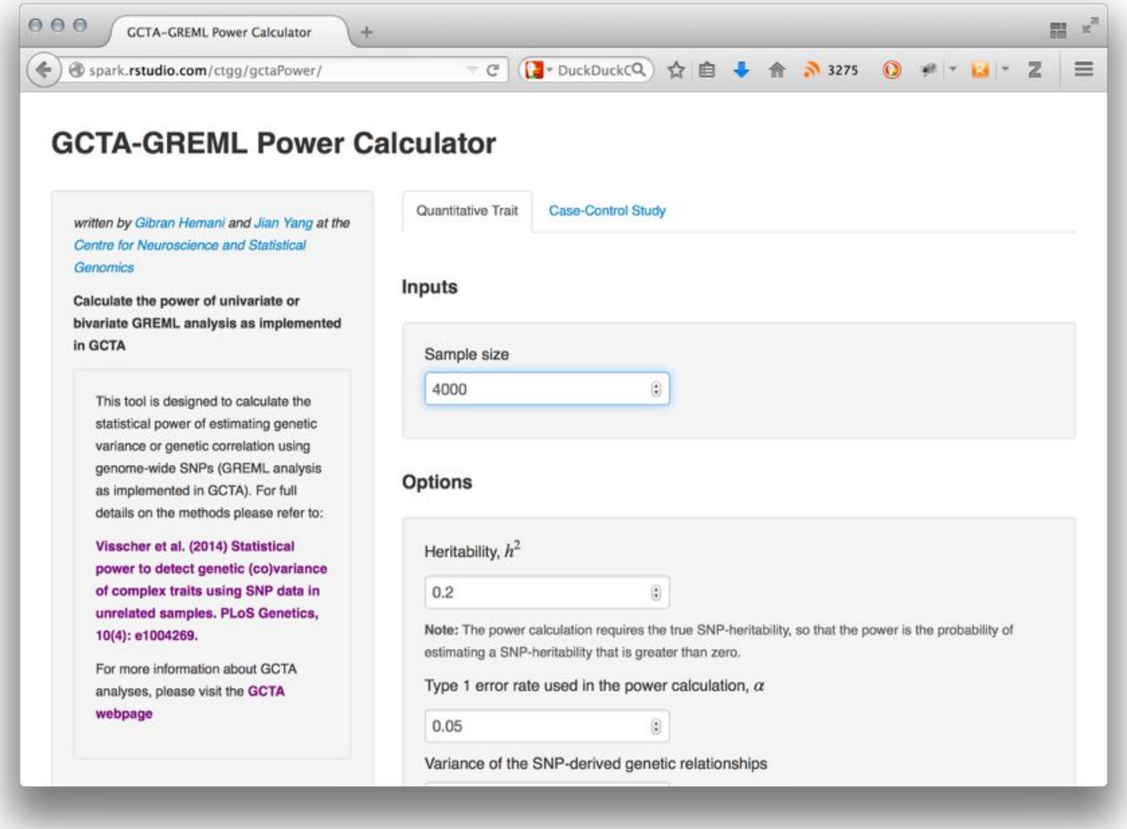
additive model,
freq=0.5, OR=1.5,
N~2 100

Statistical Power

Infinitesimal hypothesis

Statistical Power

Infinitesimal hypothesis
→ GCTA power calculator
(http://spark.rstudio.com/ctgg/gctaPower)



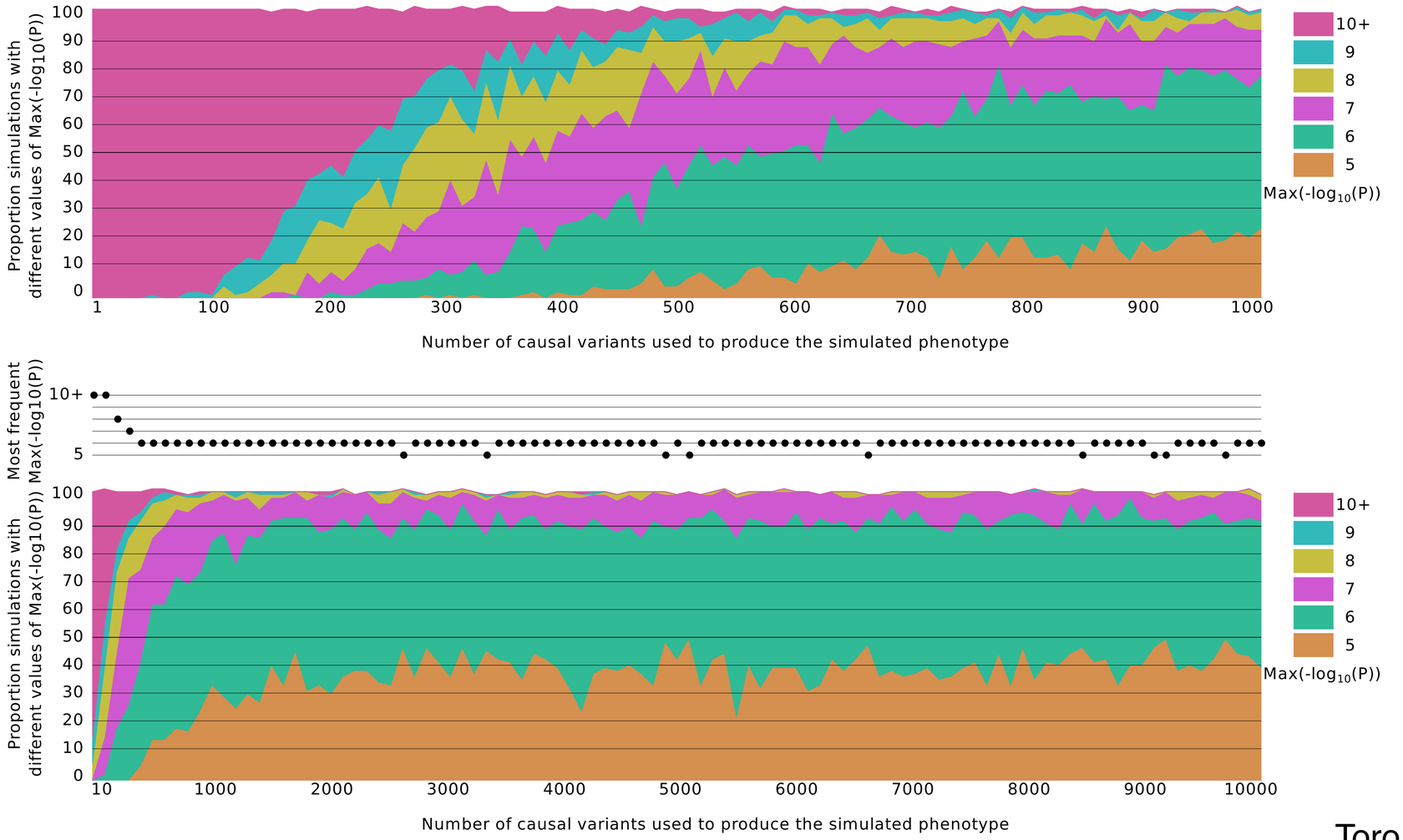
h^2=0.5, **N~1 770**
h^2=0.2, **N~4 450**

Statistical Power

Infinitesimal hypothesis, GWAS

Statistical Power

Infinitesimal hypothesis, GWAS



Toro et al, Mol Psych (2014)

Statistical Power

Infinitesimal hypothesis, GWAS

Kemper & Goddard, Hum Mol Genet (2012)

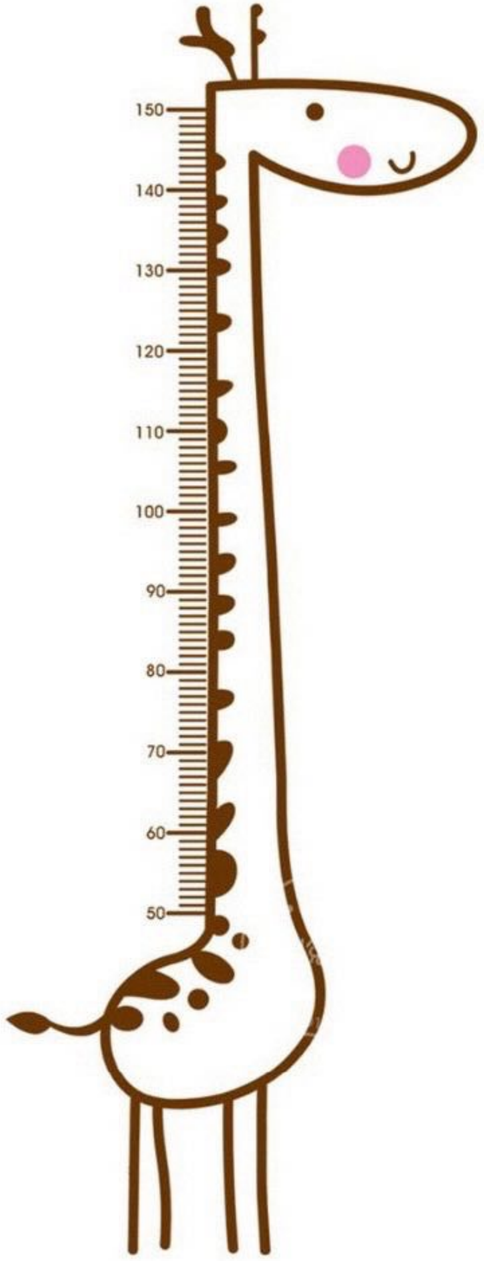
Statistical Power

Infinitesimal hypothesis, GWAS

Effective population size (Wright, 1938): “minimum number of individuals that would be required to show the same amount of dispersion of allele frequencies as the population under consideration”

For cattle, the effective population is N_e~100
⇒ **N~2 500** animals required to predict *genetic value* accurately.

For humans, the effective population is N_e~10 000–15 000
⇒ more than **N~145 000** subjects required to reach a similar accuracy.



The study of height tells us about the size of the cohorts that may be required to study quantitative traits in humans

Many sequence variants affecting diversity of adult human height

Daniel F Gudbjartsson[1], G Bragi Walters[1], Gudmar Thorleifsson[1], Hreinn Stefansson[1], Bjarni V Halldorsson[1,2], Pasha Zusmanovich[1], Patrick Sulem[1], Steinunn Thorlacius[1], Arnaldur Gylfason[1], Stacy Steinberg[1], Anna Helgadottir[1], Andres Ingason[1], Valgerdur Steinthorsdottir[1], Elinborg J Olafsdottir[3], Gudridur H Olafsdottir[3], Thorvaldur Jonsson[4], Knut Borch-Johnsen[5,6], Torben Hansen[5], Gitte Andersen[5], Torben Jorgensen[7], Oluf Pedersen[5,6], Katja K Aben[8], J Alfred Witjes[9], Dorine W Swinkels[10], Martin den Heijer[11], Barbara Franke[12], Andre L M Verbeek[13], Diane M Becker[14], Lisa R Yanek[14], Lewis C Becker[14], Laufey Tryggvadottir[3], Thorunn Rafnar[1], Jeffrey Gulcher[1], Lambertus A Kiemeney[8,9,13], Augustine Kong[1], Unnur Thorsteinsdottir[1] & Kari Stefansson[1]

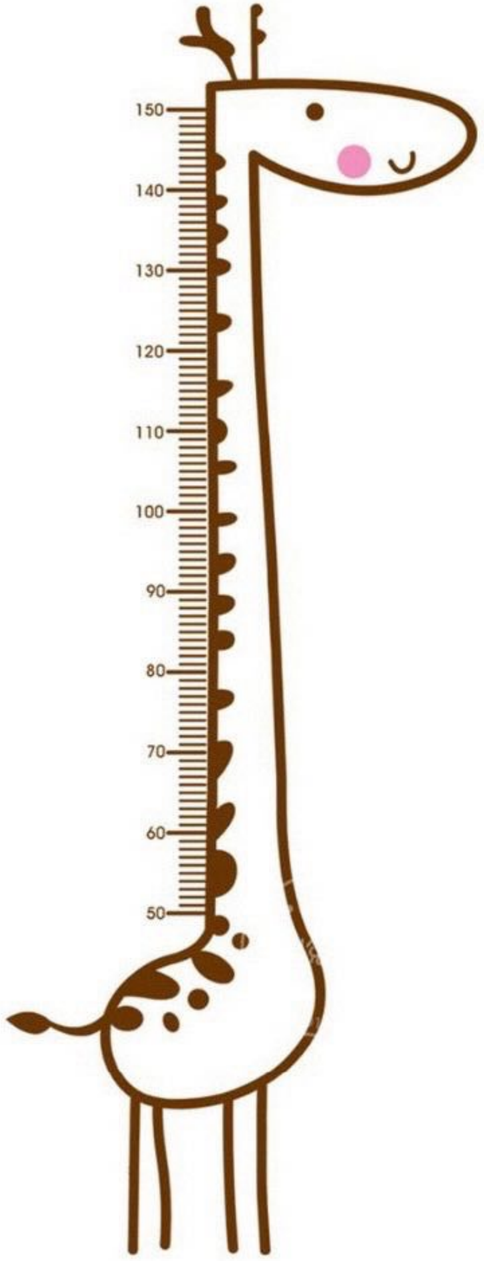
Identification of ten loci associated with height highlights new biological pathways in human growth

Guillaume Lettre[1,2], Anne U Jackson[3,25], Christian Gieger[4,5,25], Fredrick R Schumacher[6,7,25], Sonja I Berndt[8,25], Serena Sanna[3,9,25], Susana Eyheramendy[4,5], Benjamin F Voight[1,10], Johannah L Butler[2], Candace Guiducci[1], Thomas Illig[4], Rachel Hackett[1], Iris M Heid[4,5], Kevin B Jacobs[11], Valeriya Lyssenko[12], Manuela Uda[9], The Diabetes Genetics Initiative[24], FUSION[24], KORA[24], The Prostate, Lung Colorectal and Ovarian Cancer Screening Trial[24], The Nurses' Health Study[24], SardiNIA[24], Michael Boehnke[3], Stephen J Chanock[13], Leif C Groop[12,14], Frank B Hu[6,7,15], Bo Isomaa[16,17], Peter Kraft[7], Leena Peltonen[1,18,19], Veikko Salomaa[20], David Schlessinger[21], David J Hunter[1,6,7,15], Richard B Hayes[8], Gonçalo R Abecasis[3], H-Erich Wichmann[4,5], Karen L Mohlke[22] & Joel N Hirschhorn[1,2,23]

Genome-wide association analysis identifies 20 loci that influence adult height

Michael N Weedon[1,2,23], Hana Lango[1,2,23], Cecilia M Lindgren[3,4], Chris Wallace[5], David M Evans[6], Massimo Mangino[7], Rachel M Freathy[1,2], John R B Perry[1,2], Suzanne Stevens[7], Alistair S Hall[8], Nilesh J Samani[7], Beverly Shields[2], Inga Prokopenko[3,4], Martin Farrall[9], Anna Dominiczak[10], Diabetes Genetics Initiative[21], The Wellcome Trust Case Control Consortium[21], Toby Johnson[11–13], Sven Bergmann[11,12], Jacques S Beckmann[11,14], Peter Vollenweider[15], Dawn M Waterworth[16], Vincent Mooser[16], Colin N A Palmer[17], Andrew D Morris[18], Willem H Ouwehand[19,20], Cambridge GEM Consortium[22], Mark Caulfield[5],

- **N>60,000**
- **54 GW-sig. SNPs**
- **<5% of variance**



The study of height tells us about the size of the cohorts that may be required to study quantitative traits in humans

Defining the role of common variation in the genomic and biological architecture of adult human height

Using genome-wide data from 253,288 individuals, we identified 697 variants at genome-wide significance that together explained one-fifth of the heritability for adult height. By testing different numbers of variants in independent studies, we show that the most strongly associated ~2,000, ~3,700 and ~9,500 SNPs explained ~21%, ~24% and ~29% of phenotypic variance. Furthermore, all common variants together captured 60% of heritability. The 697 variants clustered in 423 loci were enriched for genes, pathways and tissue types known to be involved in growth and together implicated genes and pathways not highlighted in earlier efforts, such as signaling by fibroblast growth factors, WNT/β-catenin and chondroitin sulfate–related genes. We identified several genes and pathways not previously connected with human skeletal growth, including mTOR, osteoglycin and binding of hyaluronic acid. Our results indicate a genetic architecture for human height that is characterized by a very large but finite number (thousands) of causal variants.

- **N>250,000**
- **GW-significant 697 SNPs: 16% of variance**
- **Genome-wide SNPs: 60% of variance**
- **Probably tens of thousands of causal variants**

Further references

Sham and Purcell (2014) *Statistical power and significance testing in large-scale genetic studies, Nat Rev Genet, doi:10.1038/nrg3706*

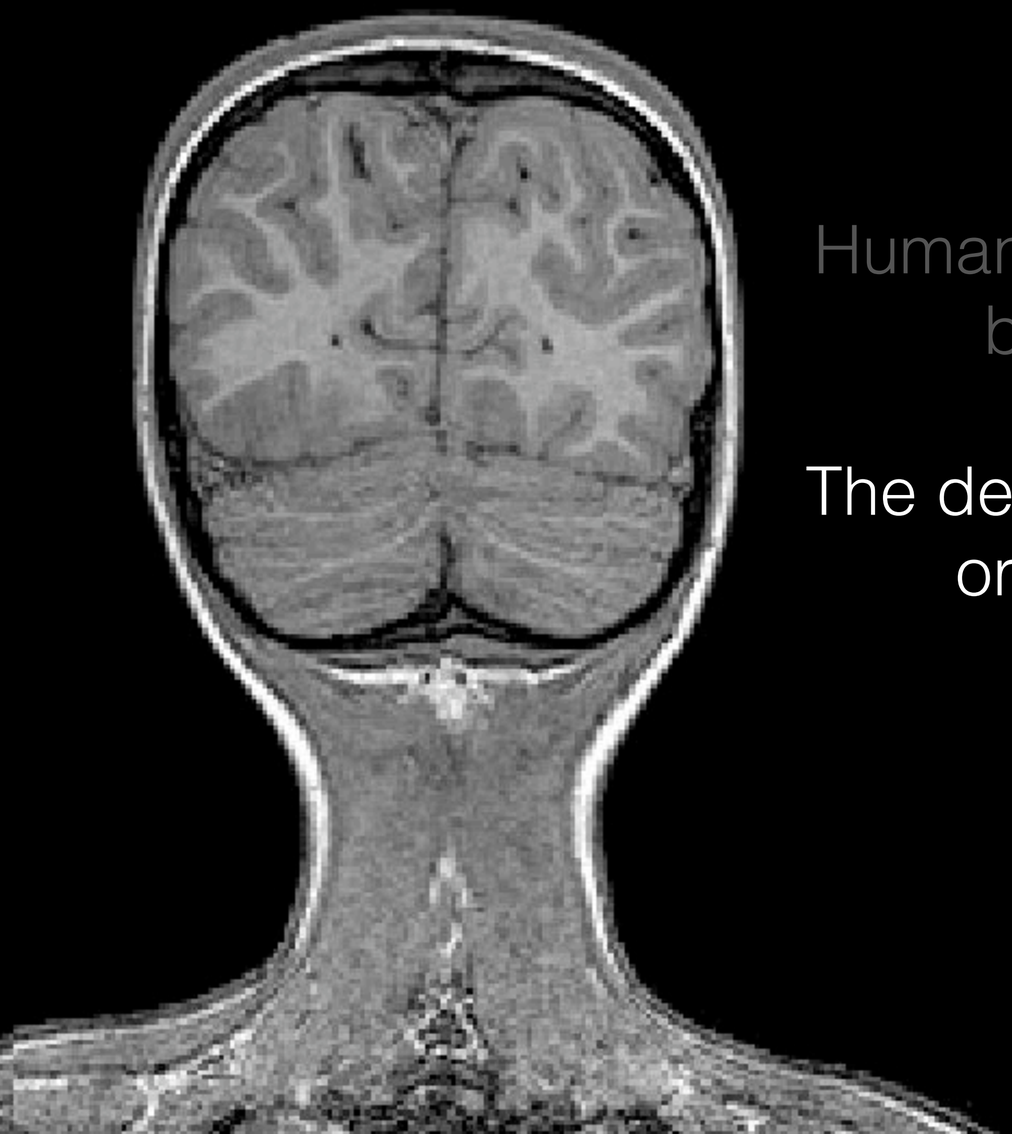
Visscher et al (2014) *Statistical Power to Detect Genetic (Co)Variance of Complex Traits Using SNP Data in Unrelated Samples, PLoS Genet, doi:10.1371/journal.pgen.1004269*

Kemper and Goddard (2012) *Understanding and predicting complex traits: knowledge from cattle, Hum Mol Genet, doi:10.1093/hmg/dds332*

Bulik-Sulivan et al (2015) *LD Score regression distinguishes confounding from polygenicity in genome-wide association studies, Nat Genet, doi: 10.1038/ng.3211*

Gusev et al (2014) *Regulatory variants explain much more heritability than coding variants across 11 common diseases, BiorXiv, doi: 10.1101/004309*

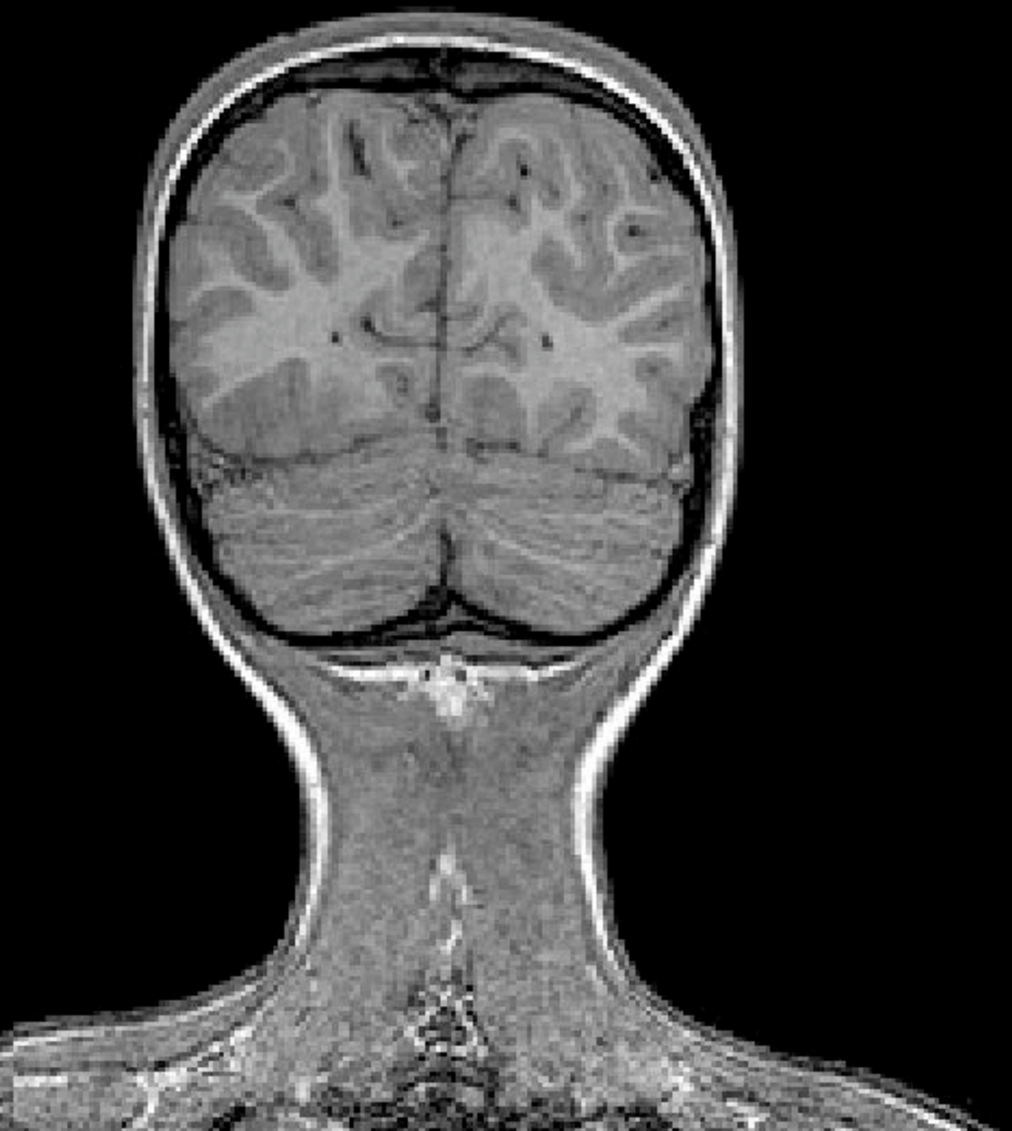
Yang et al (2013) *Ubiquitous Polygenicity of Human Complex Traits: Genome-Wide Analysis of 49 Traits in Koreans. PLoS Genet, doi: 10.1371/journal.pgen.1003355*



Human brain diversity and pathology may be encoded by **tens of thousands** of genomic variants.

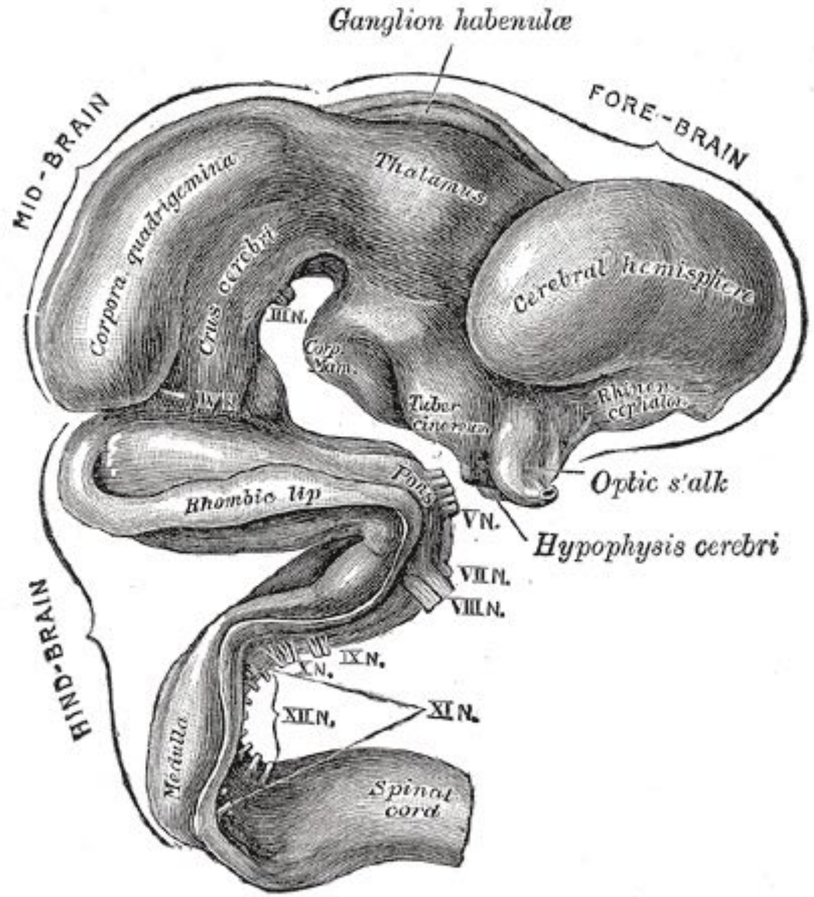
The detection of these variants may require **thousands** or even **hundreds of thousands** of subjects.

This puts strong limitations on the type of imaging sequence and the type of analyses we can use, which have to be **accurate**, **reproducible** and as much as possible **automatic**.



2.
Neuroimaging phenotypes

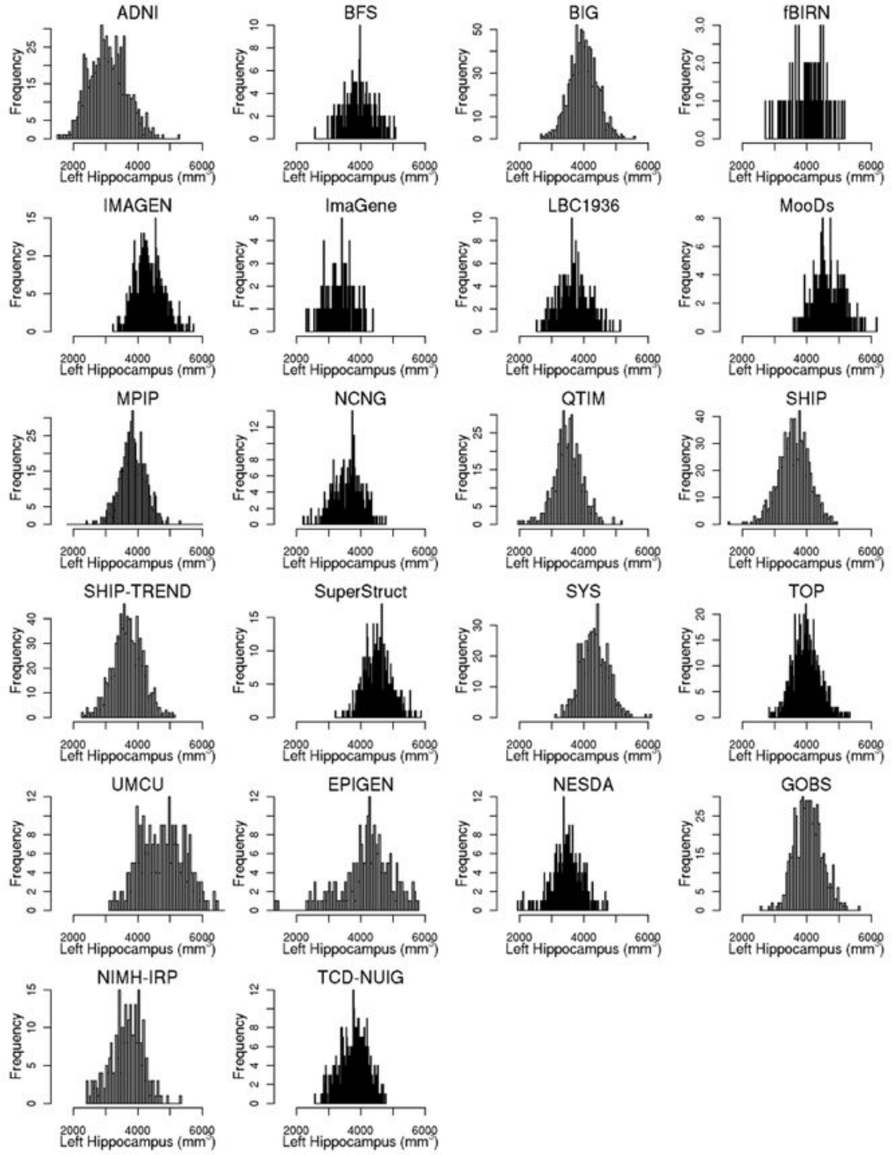
early development



Brain region	h^2 (95% CI)	Reference
ICV	79% (52-87)	Kremen et al (2010)
BV	70% (34-81)	Yoon et al (2011)
Hippocampus	76% (66-83)	den Braber et al (2013)
Thalamus	81% (74-85)	den Braber et al (2013)
Caudate nucleus	87% (82-91)	den Braber et al (2013)
Pallidum	70% (56-80)	den Braber et al (2013)
Putamen	85% (56-90)	den Braber et al (2013)
Amygdala	67% (57-76)	den Braber et al (2013)
Accumbens	67% (56-75)	den Braber et al (2013)

Kremen et al: N=474 (202 twin pairs, 70 unpaired)
Yoon et al: N=184 (57 MZ, 35 DZ, infants)
den Braber et al: N=528 (176 MZ pairs, 88 DZ pairs)

"Identification of common variants associated with human hippocampal and intracranial volumes", Stein et al, Nat Genet 2012



Hippocampal volume

TESC (rs7294919) (regulation of intracellular pH, cell volume and cytoskeletal organization)

0008285 negative regulation of cell proliferation
0010628 positive regulation of gene expression
0033628 regulation of cell adhesion mediated by integrin
0045654 positive regulation of megakaryocyte differentiation
0043193 positive regulation of gene-specific transcription

Intracranial volume

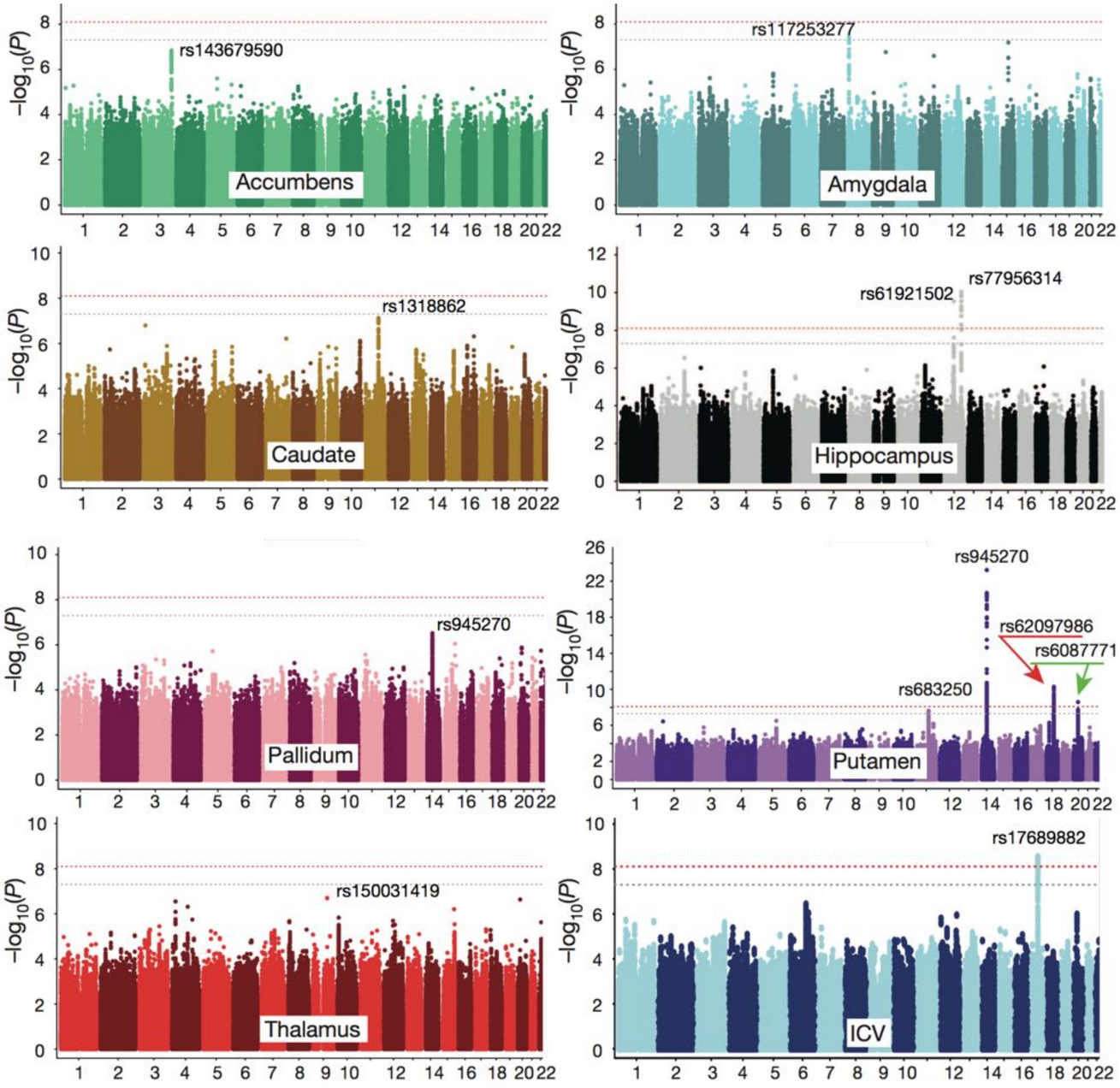
HMGA2 (rs10784502) (already associated with height)

0051301 cell division
0007049 cell cycle
0006325 chromatin organization
0007275 multicellular organismal development
0007067 mitosis
0006355 regulation of transcription, DNA-dependent
0040008 regulation of growth

N: 7,795
(authors N: 209, +consortia N~1 200, i.e., 8 subj/auth)

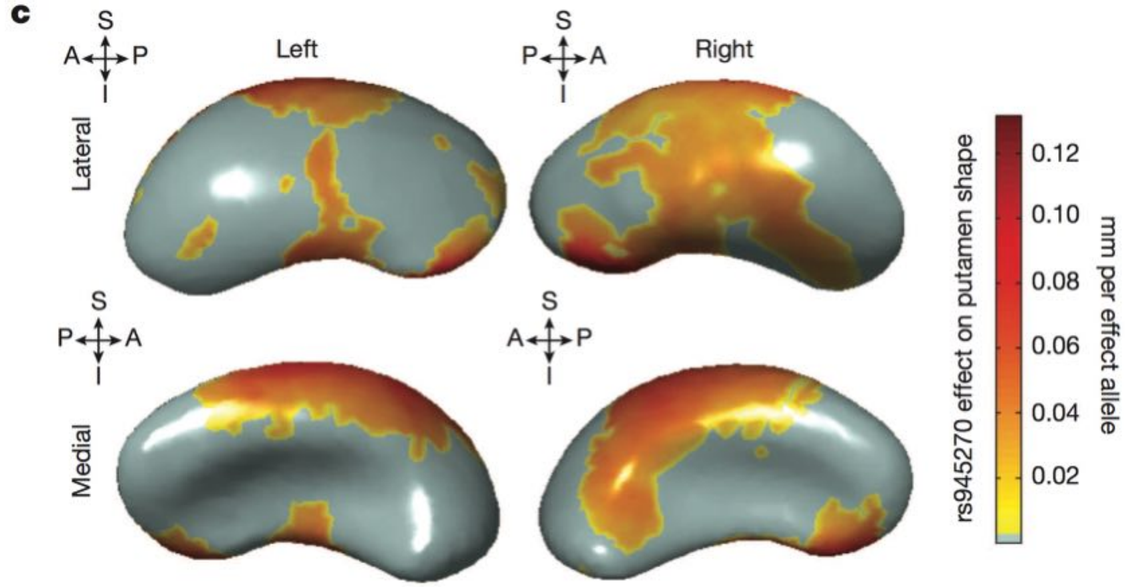
Also: Ikram et al, Nature Genetics 2012, Taal et al, Nature Genetics 2012

"Common genetic variants influence human subcortical brain structures", Hibar et al, Nature 2015

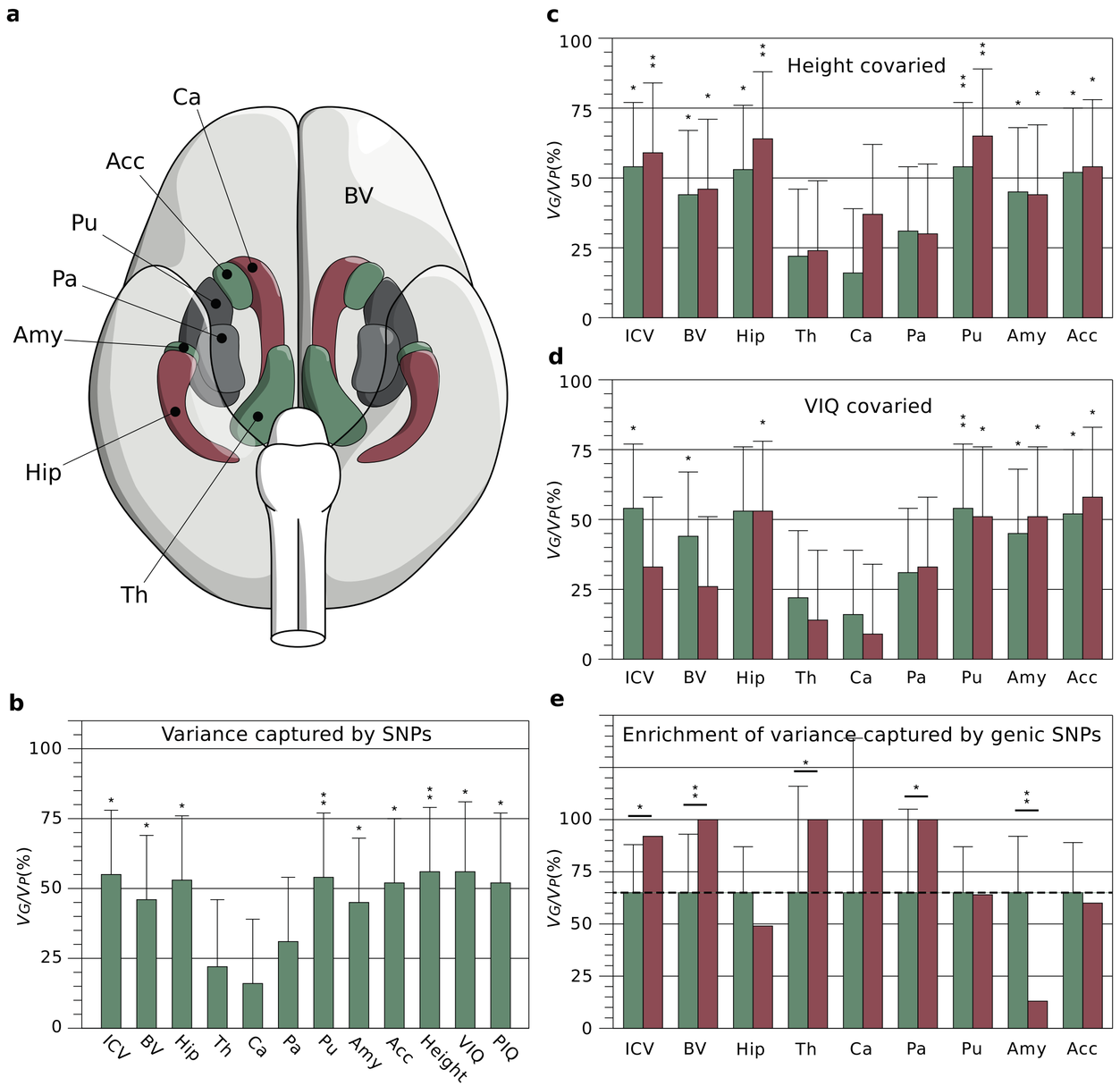


N: 30,717 (50 cohorts)

- validation of SNPs previously found for Hipp & ICV
- New variants for Putamen and Caudate Nucleus
- Significant SNPs explain 0.17-0.52% of the variance
- Validation in IMAGEN: shape analysis

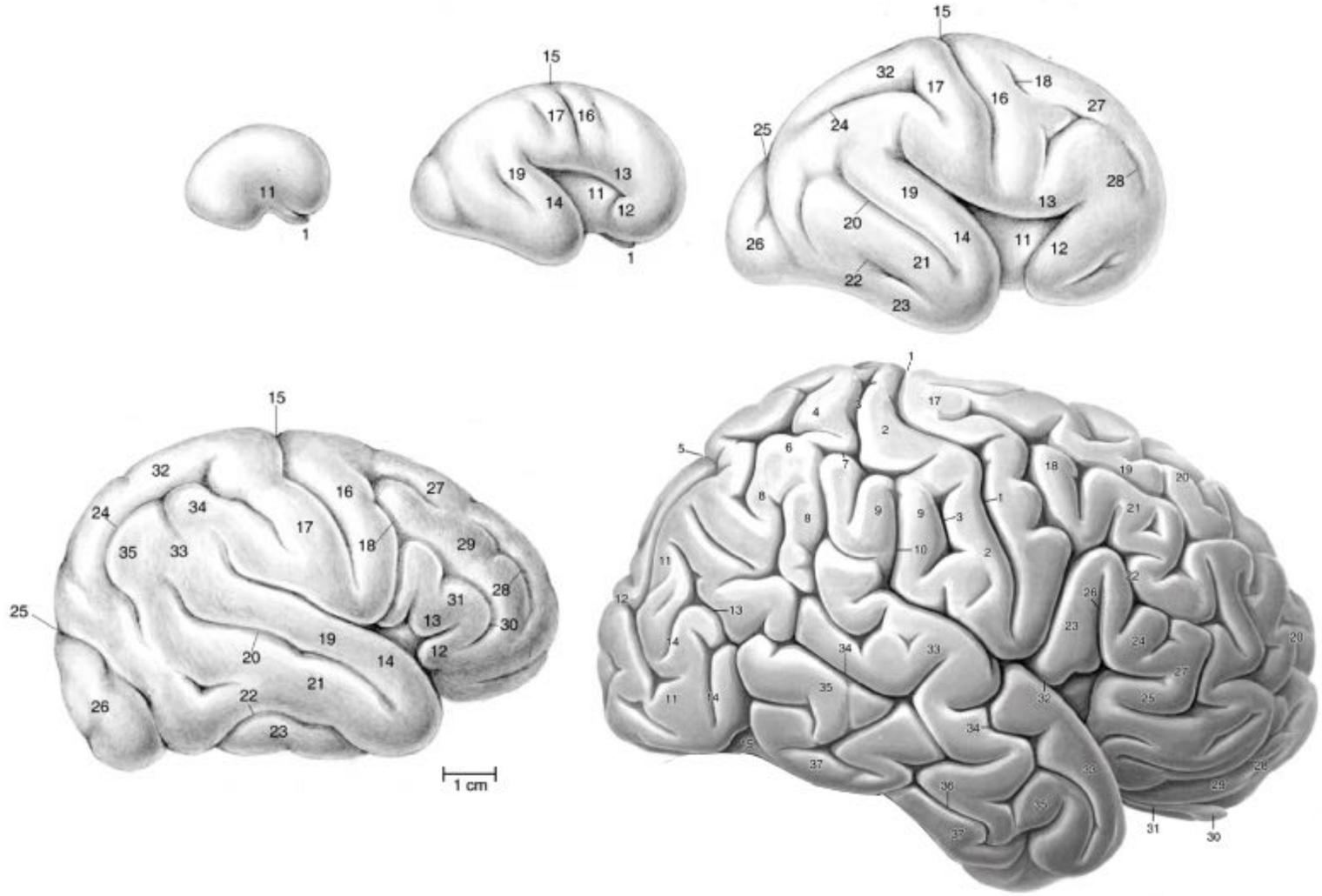


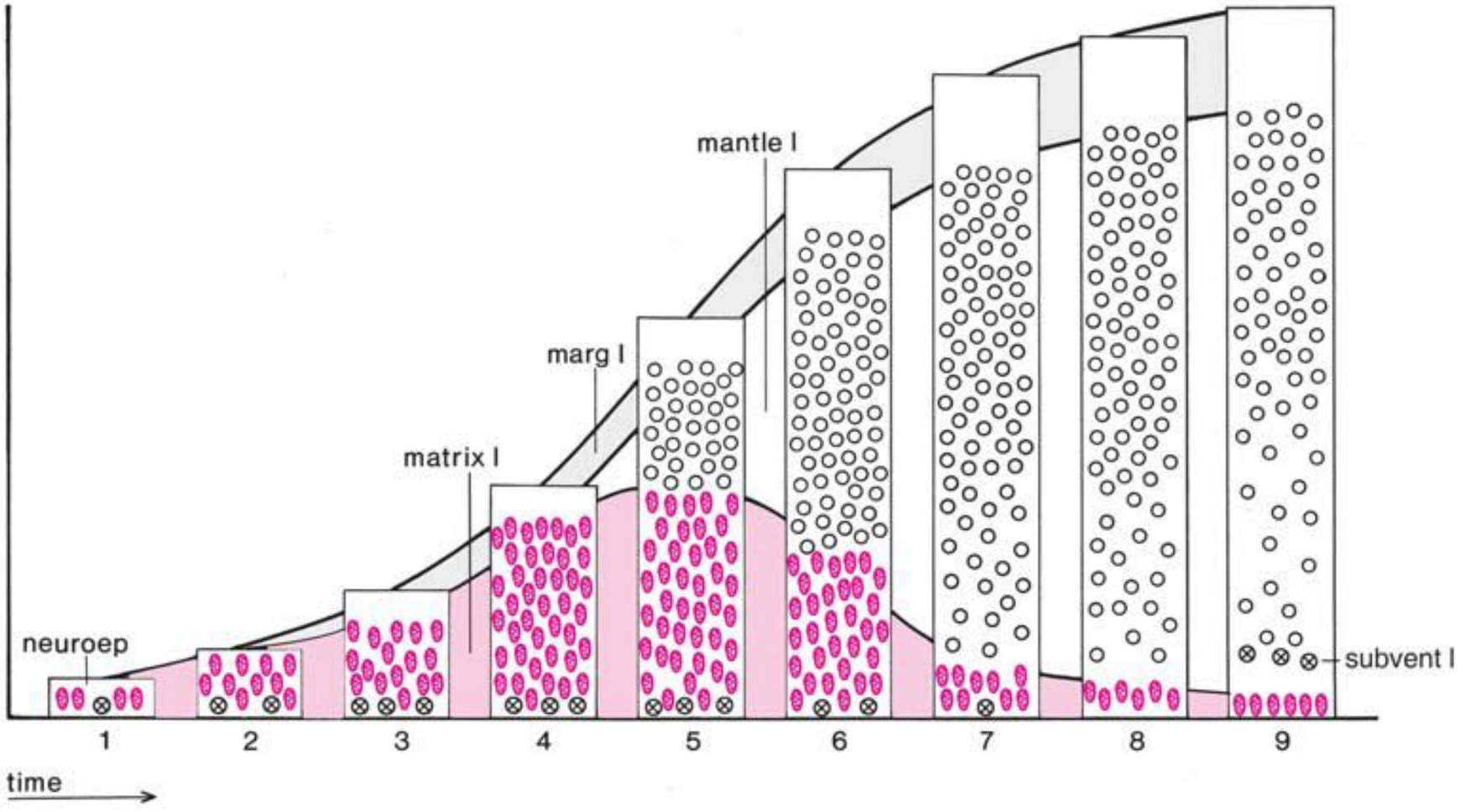
"Genomic Architecture of Human Neuroanatomical Diversity", Toro et al, Mol Psych 2014

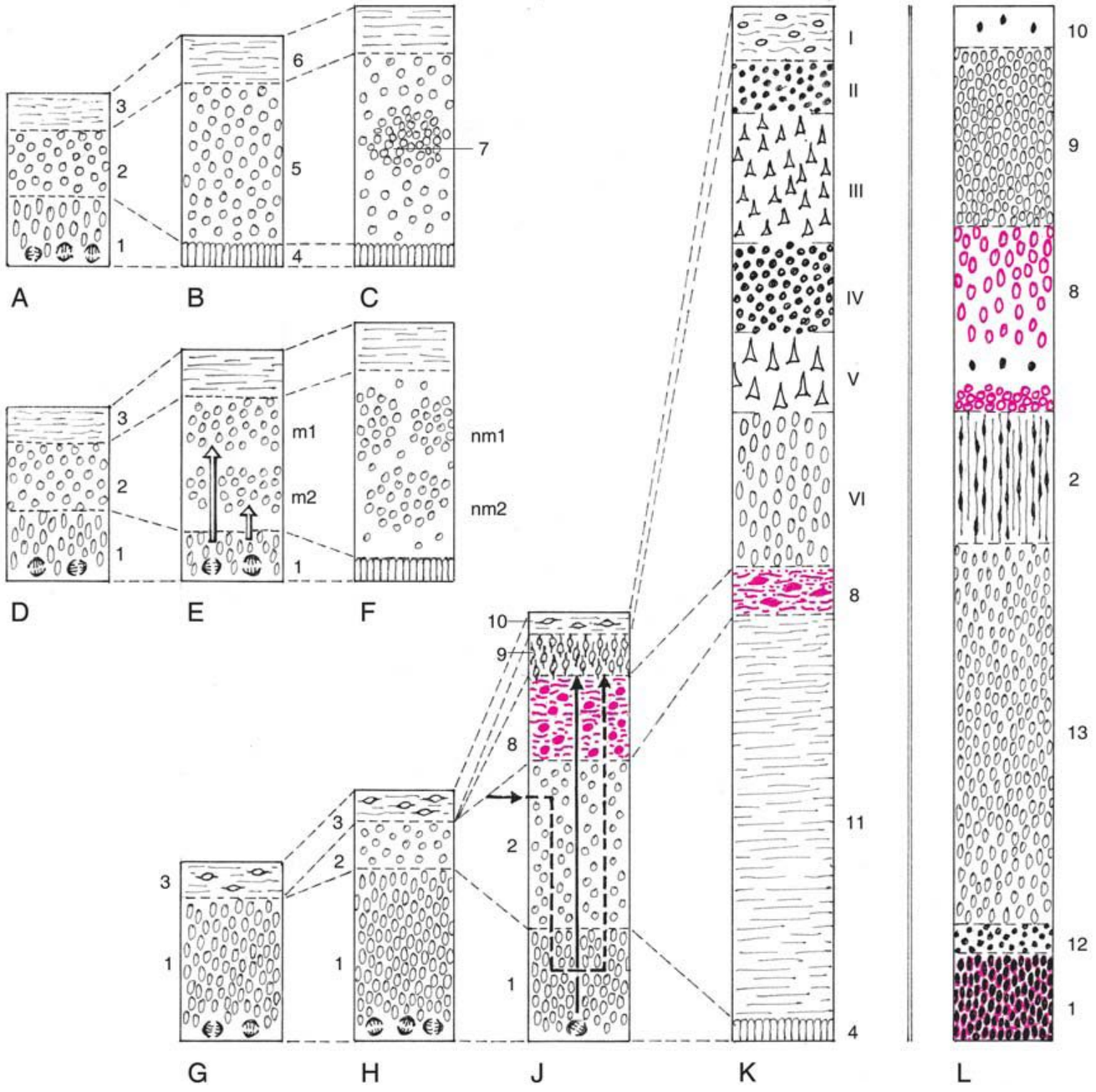


- IMAGEN cohort: **N=1,765**, 14-15 years old.
- **~50%** of the variance neuroanatomical variance is captured by the additive effect of **100s and possibly 1000s** of Single Nucleotide polymorphisms (SNPs), each of **small effect**.
- **~90%** of this variance is captured by SNPs **within genes and close regulatory regions**.
- The genetic bases of neuroanatomical diversity are **strongly correlated with those of verbal IQ**, but **less so with those of body size** (height).
- **Very large populations** (N>100k) will be required to capture any sizeable proportion of neuroanatomical diversity through GWAS.

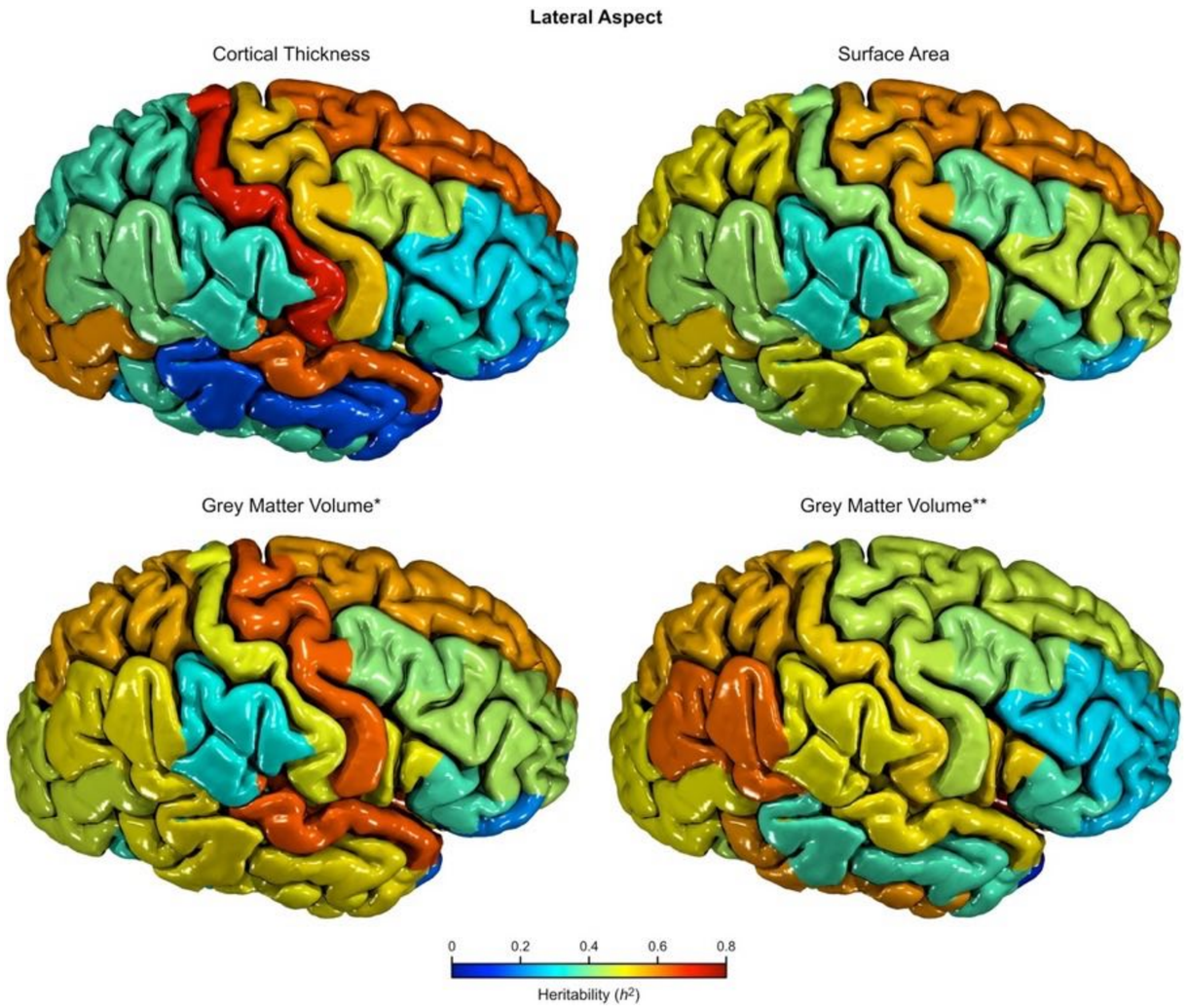
cortical development




mantle l
marg l
matrix l
neuroep
subvent l
1
2
3
4
5
6
7
8
9
time


A
B
C
D
E
F
G
H
J
K
L
m1
m2
nm1
nm2
I
II
III
IV
V
VI

***"Cortical thickness or gray matter volume? The importance of selecting the phenotype for imaging genetic studies"*, Winkler et al, NeuroImage 2009**



- N: 486 (extended pedigree)

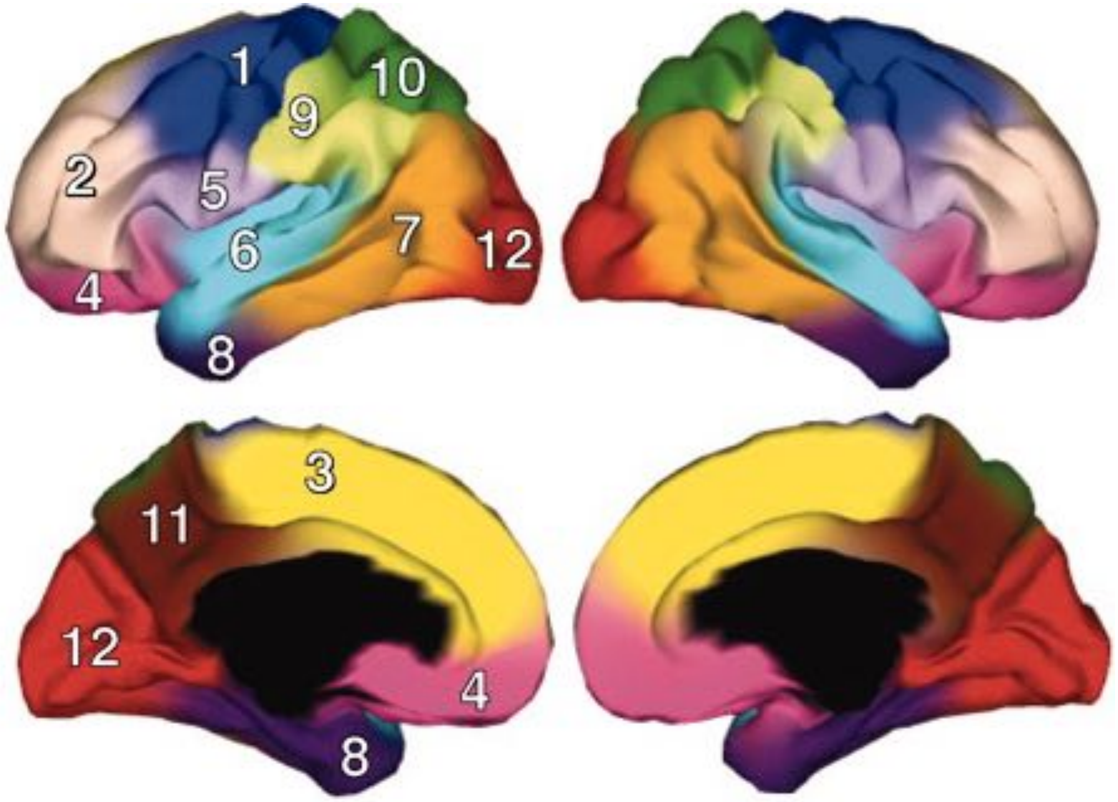
h^2(Brain Volume) = 70%
h^2(Surface) = 70%
h^2(Thickness) = 69%
h^2(GM surface-based) = 72%
h^2(GM voxel-based) = 67%

- The heritability of surface, thickness and grey matter volume were high.

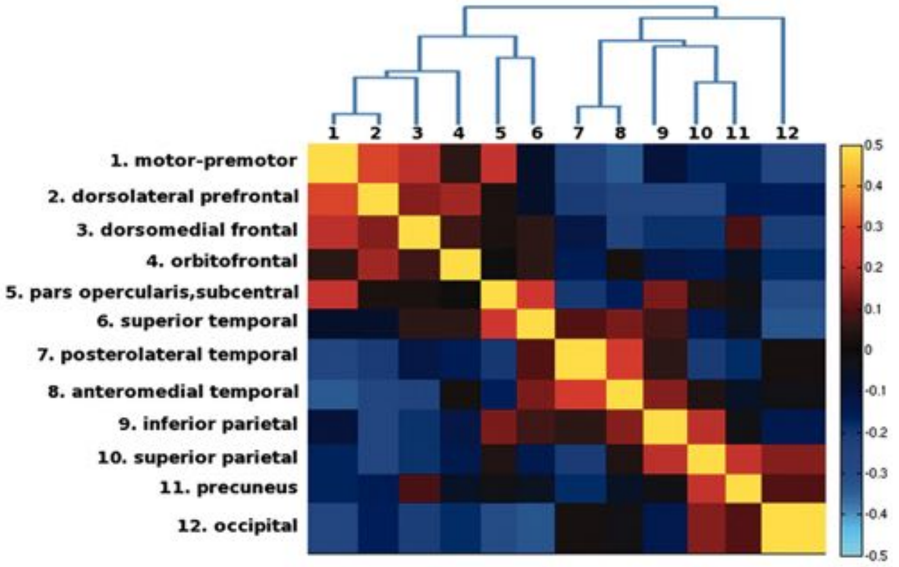
- The low genetic correlation between the additive genetic factors of surface and thickness (r_g=-0.15) suggests that different genetic factors are involved in the their development.

Also:
"Distinct genetic influences on cortical surface and cortical thickness", Panizzoni et al, Cereb Cortex 2009
"Cortical thickness is influenced by regionally specific genetic factors", Rimol et al, Biol Psychiatry 2010
"Heritable changes in regional cortical thickness with age", Chouinard-Decorte et al, Brain Imaging Behav 2014

***"Hierarchical genetic organization of human cortical surface area",* Chen et al, Science 2012**

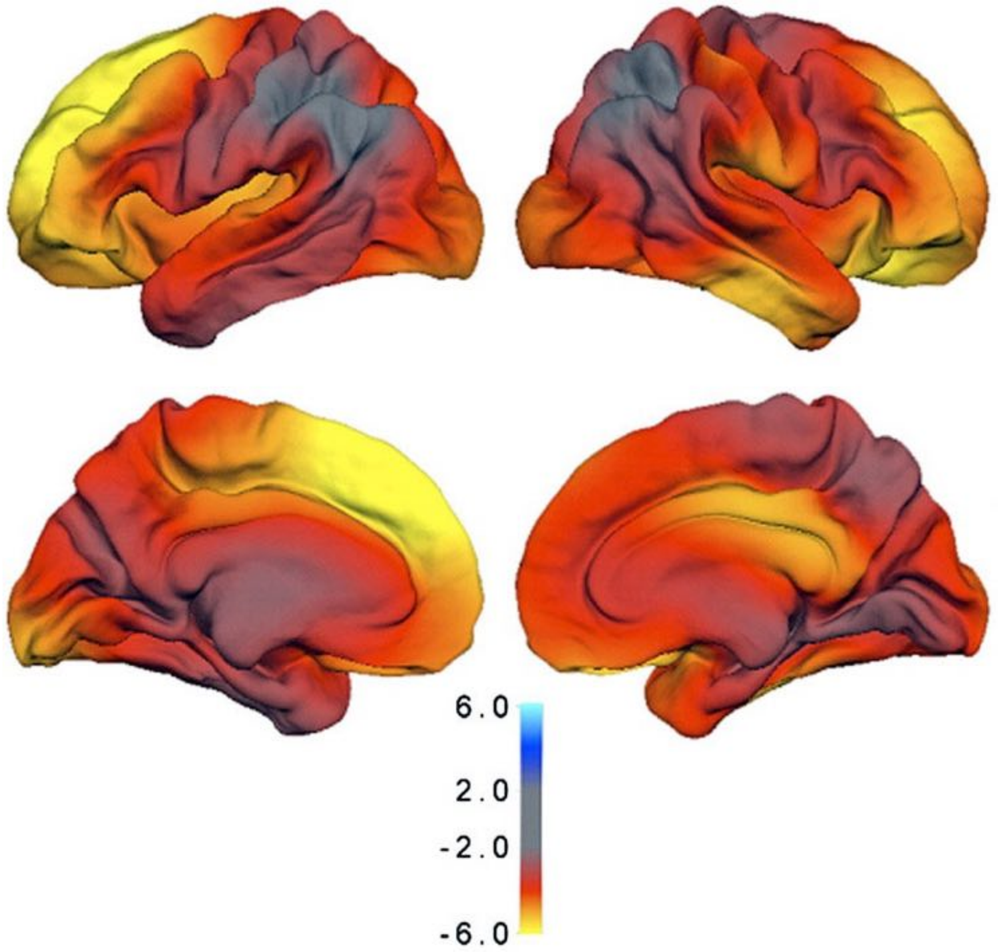


- N: 406 (110 MZ, 93 DZ)
- Phenotype: local cortical area expansion
- Genetic-correlation-based parcellation
- The genetic organisation of cortical area was hierarchical, modular, and predominantly bilaterally symmetric



Also:
"Genetic Influences on Cortical Regionalization in the Human Brain", Chen et al, Neuron 2011

***"Sex-dependent association of common microcephaly genes with brain structure",* Rimol et al, PNAS 2010**



CDK5RAP2/MCPH3 (rs4836817, rs10818453, rs4836819, rs4836820, rs7859743, rs2297453, rs2282168, rs1888893, rs914592, rs914593)

0045664 regulation of neurone differentiation (11 genes)
0007420 brain development (91 genes)

MCPH1 (rs2816514, rs2816517, rs11779303, rs11779303)

[no biol. proc. in GO

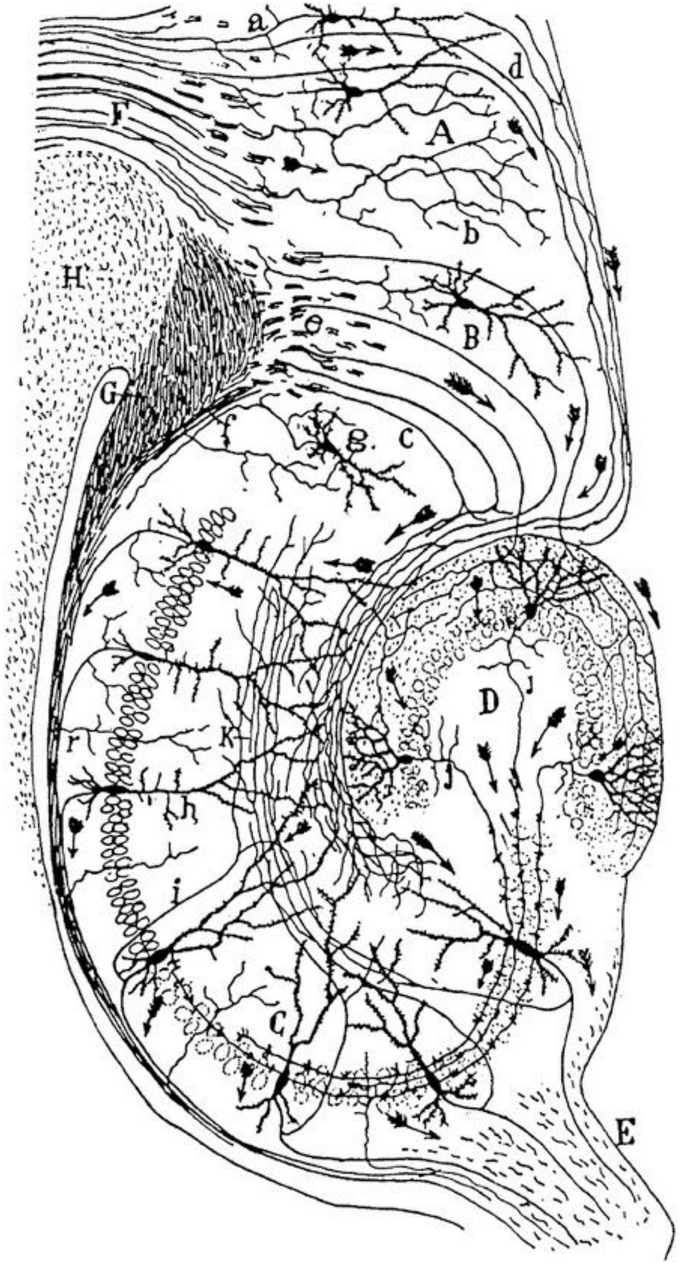
ASPM (rs10922168)

0007049 cell cycle (443 genes)
0007067 mitosis (171 genes)
0051301 cell division (221)

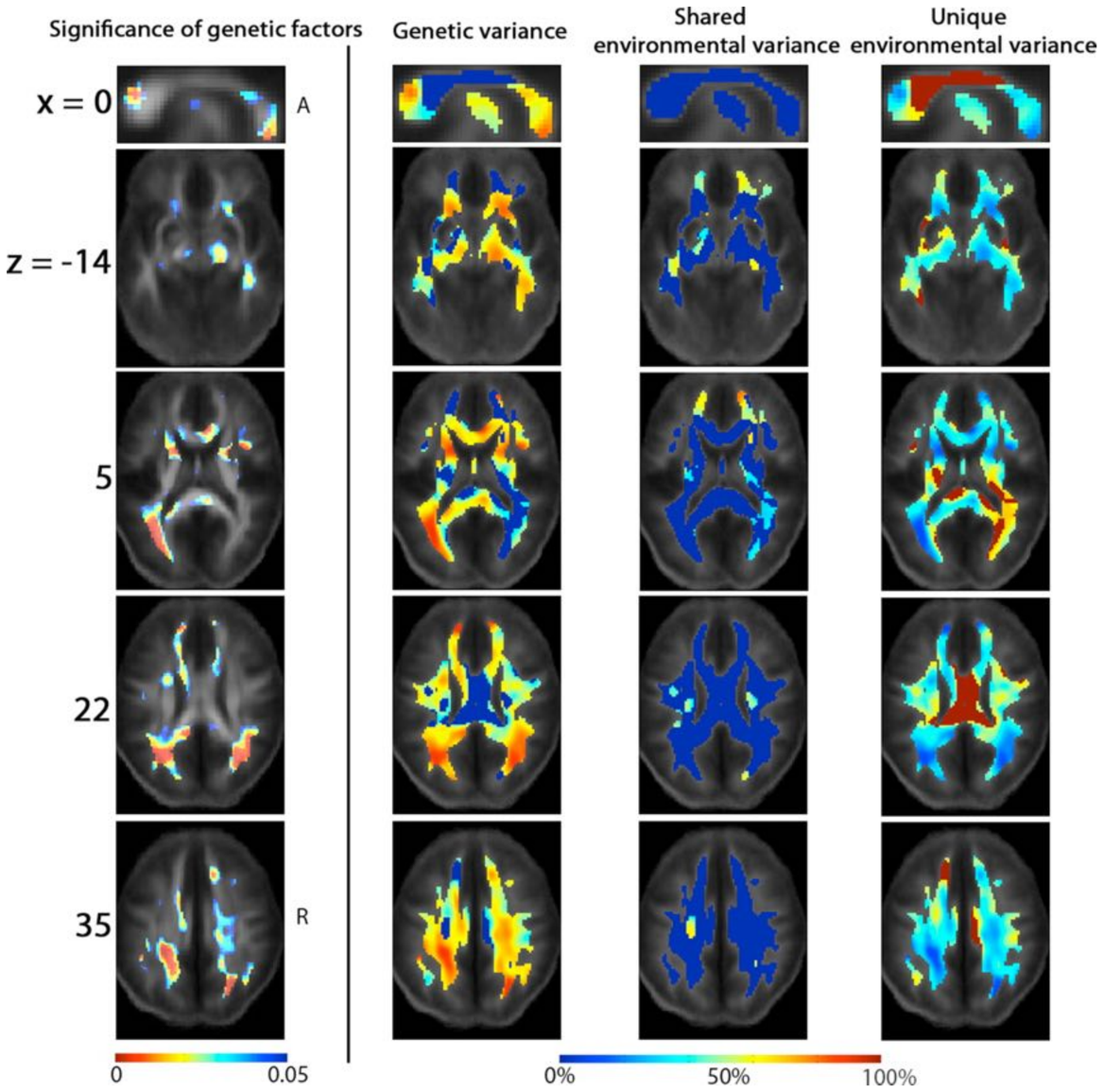
- Phenotype: local cortical area expansion
- N=287, replication N=656
- MCPH1, ASPM: Significant in females
- MCPH3: Significant in males

Accurate cortical reconstructions can be obtained automatically using Freesurfer (http://www.freesurfer.net). Several measurements of cortical geometry can be obtained using Mindboggle (http://www.mindboggle.org).

connectivity



"Genetics of brain fiber architecture and intellectual performance", Chiang et al, J Neurosci 2009

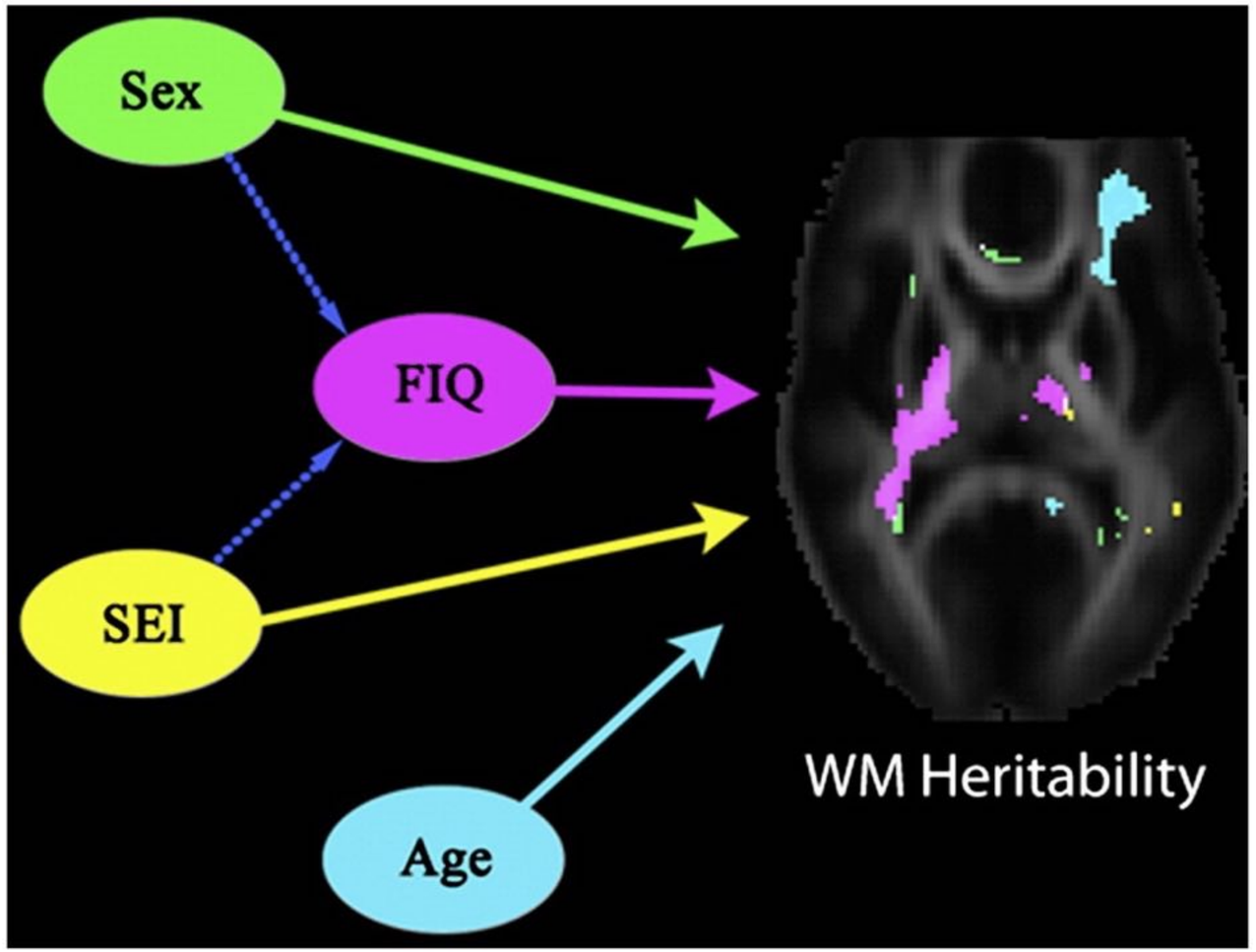


N: 92 (MZ=2*23, DZ=2*23)
h^2(FA) values from 55% (Frontal left) to 85% (Parietal left)

• The genetic determinants of FA seem to be shared with those of IQ.

Also:
"Genetic influences on brain asymmetry: a DTI study of 374 twins and siblings", Janhashad et al, NeuroImage 2010

"Genetics of white matter development: A DTI study of 705 twins and their siblings aged 12 to 29", **Chiang et al, Neuroimage 2011**

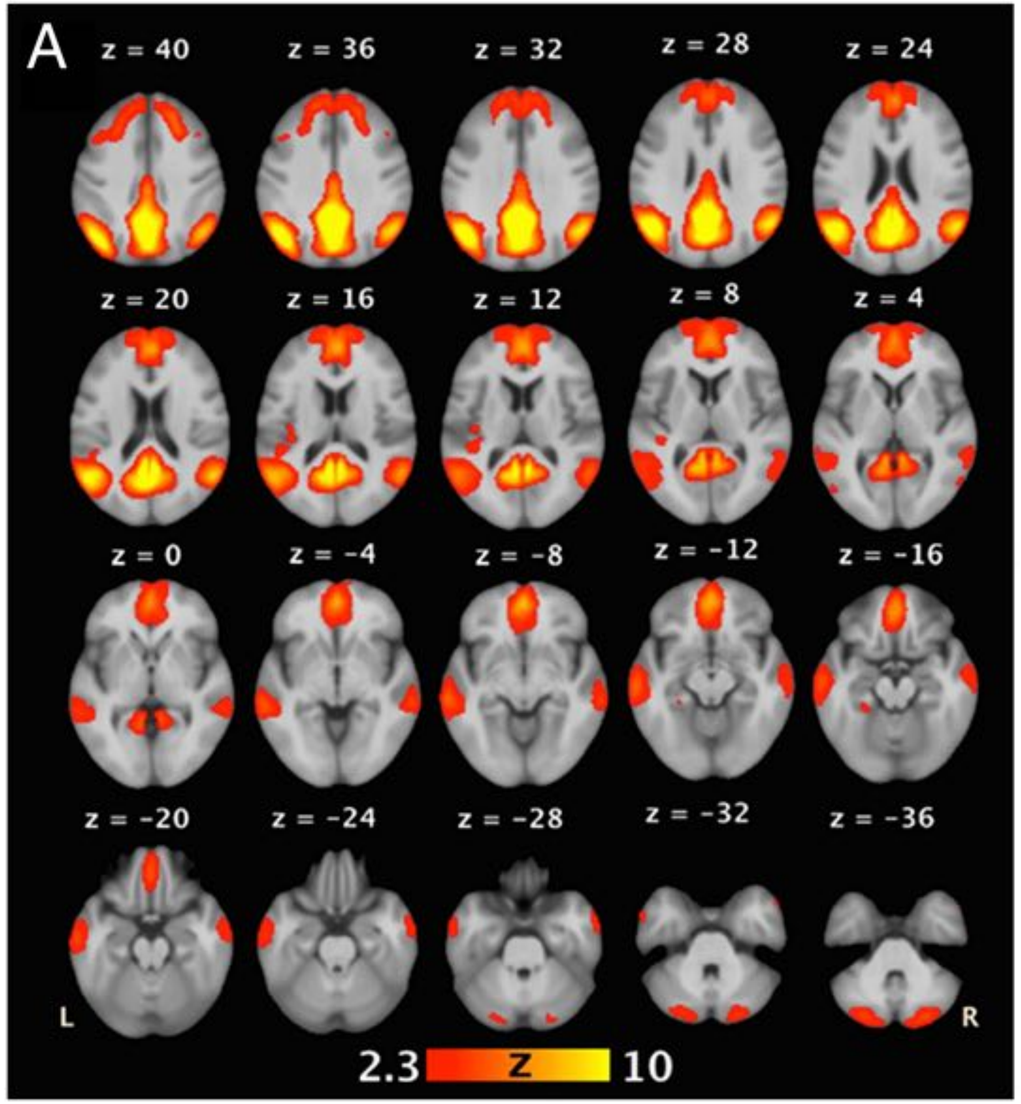


- N: 705 (119 MZ, 152 DZ, 5 TZ, + sibs)
- In adolescents: h^2(FA)=70–80%, in adults: h^2(FA)=30–40%
- h^2(FA) larger in males than in females
- h^2(FA) is modulated by socioeconomic status (larger in some regions, smaller in others)

Also: Jahanshad, N. et al. Multi-site genetic analysis of diffusion images and voxelwise heritability analysis: a pilot project of the ENIGMA-DTI working group. Neuroimage 81, 455–469 (2013).

Enigma is working to develop reliable methods for DTI genetics analyses: http://enigma.ini.usc.edu/protocols/dti-protocols/

"Genetic control over the resting brain", Glahn et al, PNAS 2010

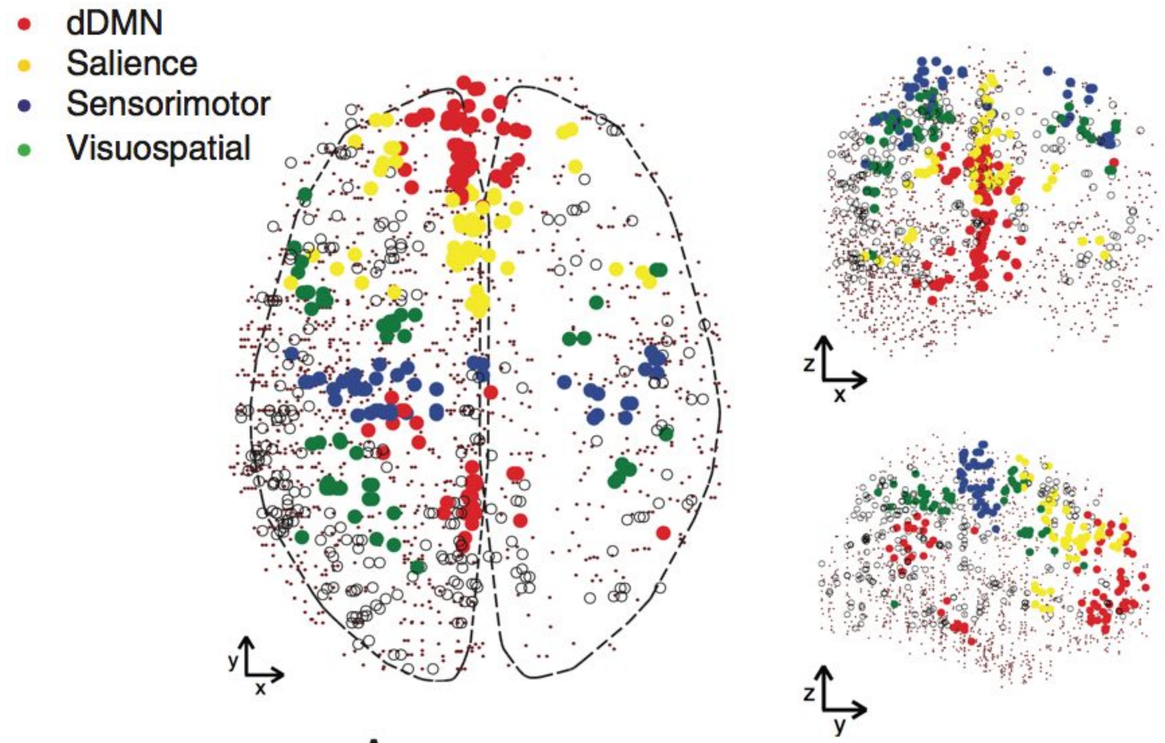


- N: 333 (extended pedigree)
- h^2(Funct. Conn) = 42%
- h^2(GM density) = 32%
- r_g = 0.07

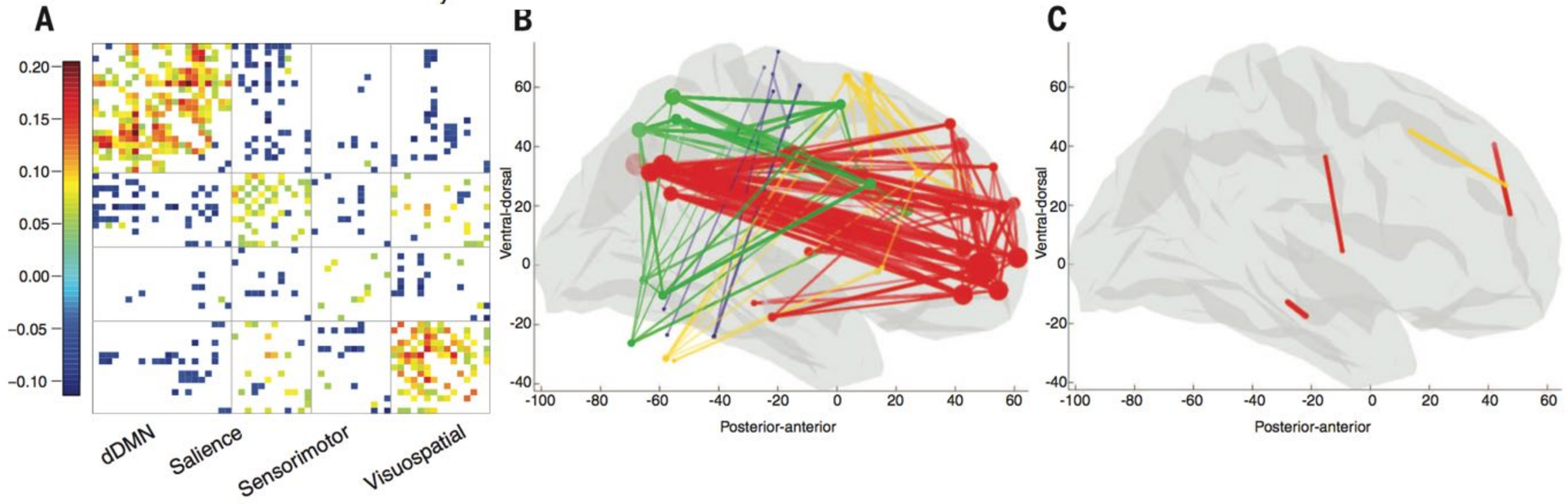
- Genes involved in functional connectivity are different from those involved in brain anatomy

Scripts for multi-centric analyses of functional connectivity have are available from the 1000 functional connectomes project: http://www.nitrc.org/projects/fcon_1000/

"Correlated gene expression supports synchronous activity in brain networks", Richiardi et al, Science 2015



- rfMRI networks extracted from a group 15 subjects
- Gene expression data from Allen Institute of Brain Research, 6 subjects
- Human validation in 259 subjects from IMAGEN
- Correspondance with axonal connectivity in mice

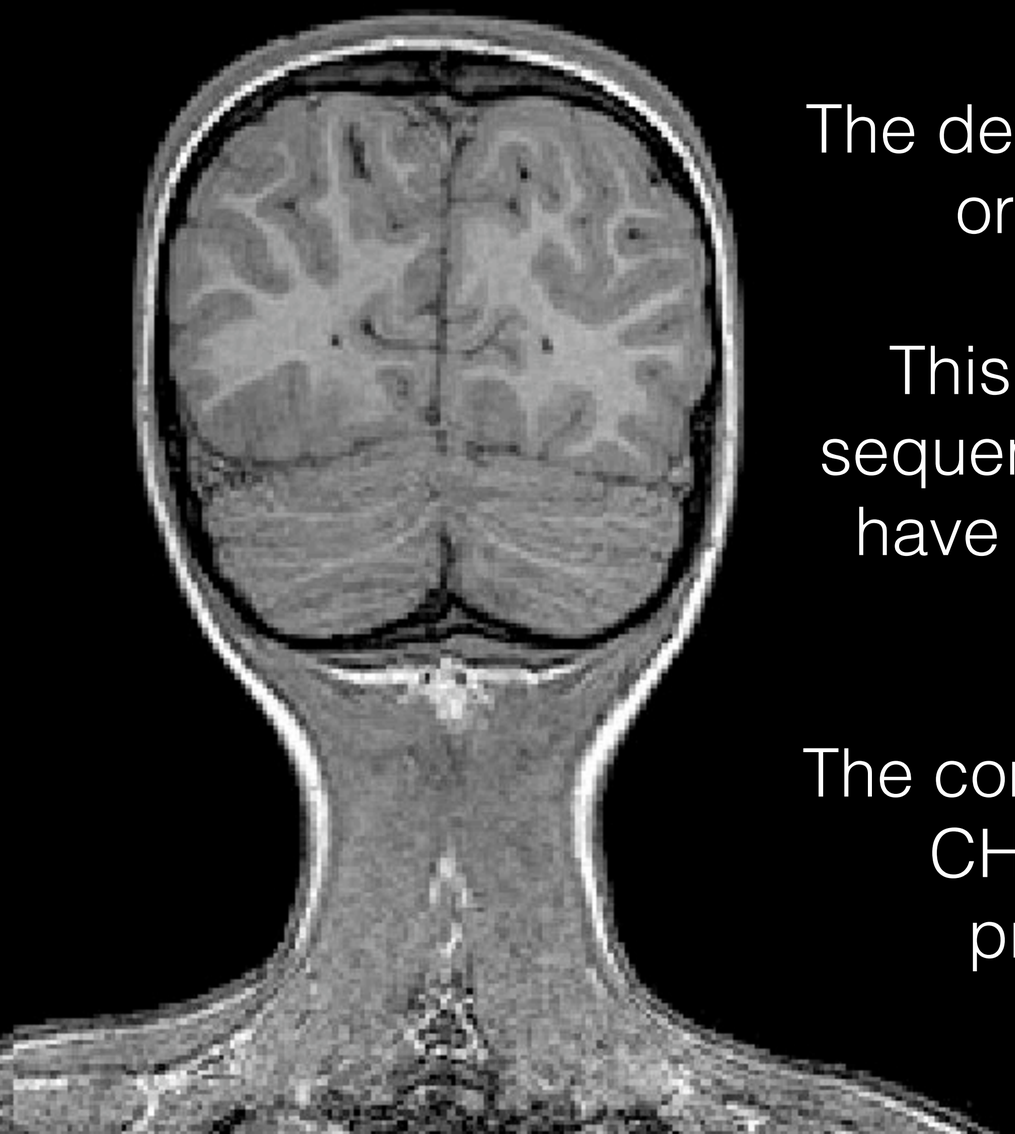


Further references

Thompson et al (2013) *The ENIGMA Consortium: Large-scale Collaborative Analyses of Neuroimaging and Genetic Data, Brain Imaging and Behav, doi: 10.1007/s11682-013-9269-5*

Medland et al (2014) *Whole-genome analyses of whole-brain data: working within an expanded search space, Nat Neurosci, doi:10.1038/nn.3718*

Gratten et al (2014) *Large-scale genomics unveils the genetic architecture of psychiatric disorders, Nat Neurosci, doi: 10.1038/nn.3708*

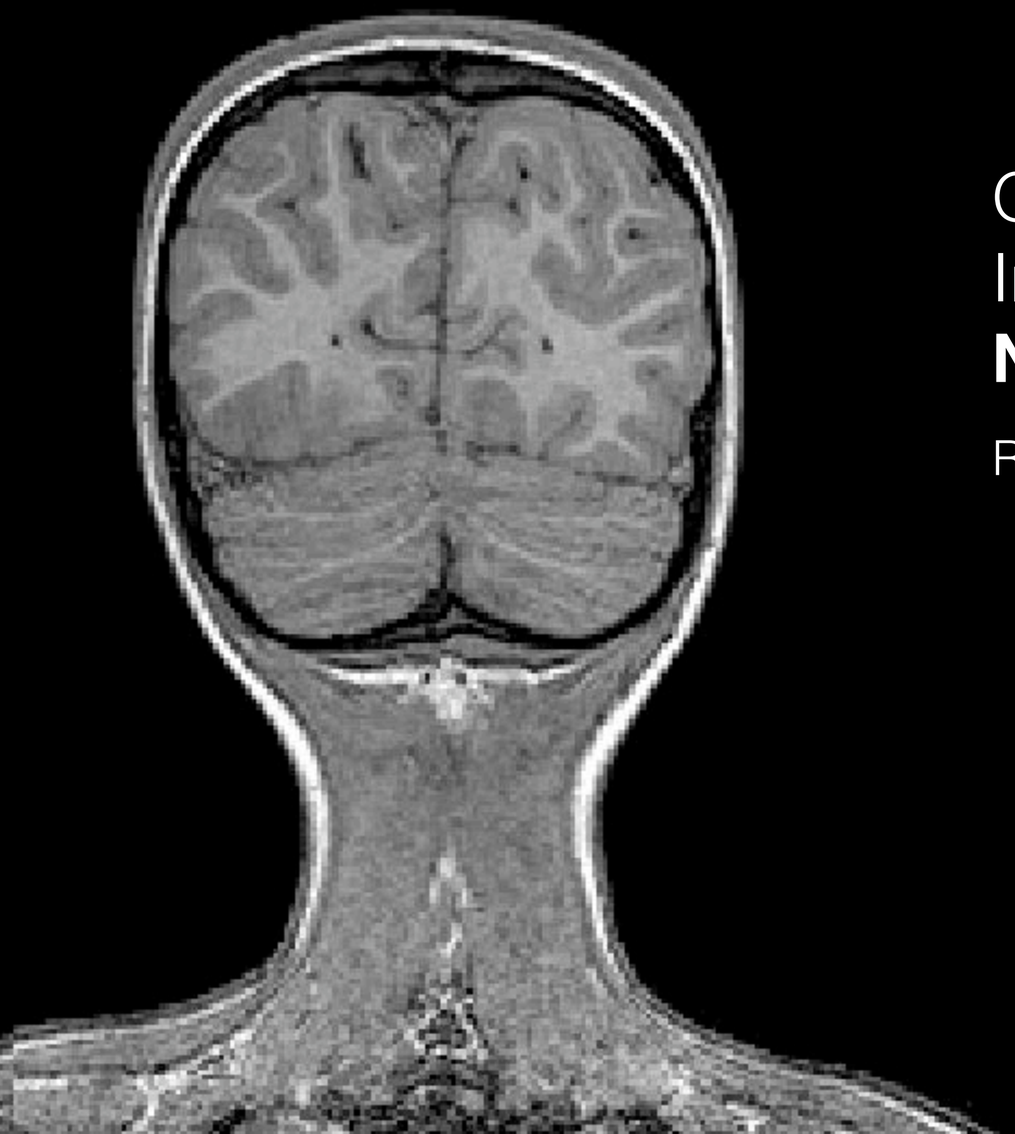


Human brain diversity and pathology may be encoded by **tens of thousands** of genomic variants.

The detection of these variants may require **thousands** or even **hundreds of thousands** of subjects.

This puts strong limitations on the type of imaging sequence and the type of analyses we can use, which have to be **accurate**, **reproducible** and as much as possible **automatic**.

The constitution of **large consortia** such as ENIGMA or CHARGE, and the development of methods for processing **multi-centric data** are essential.



OHBM 2015

Imaging Genetics Course

Neuroimaging phenotypes and heritability

ROBERTO TORO • INSTITUT PASTEUR

rto@pasteur.fr

http://neuroanatomy.github.io

Twitter: R3RT0