Private Genomes and Public Alleles

Richard Mott*[1], Christian Fischer[2], Pjotr Prins[2] and Robert W Davies[3]

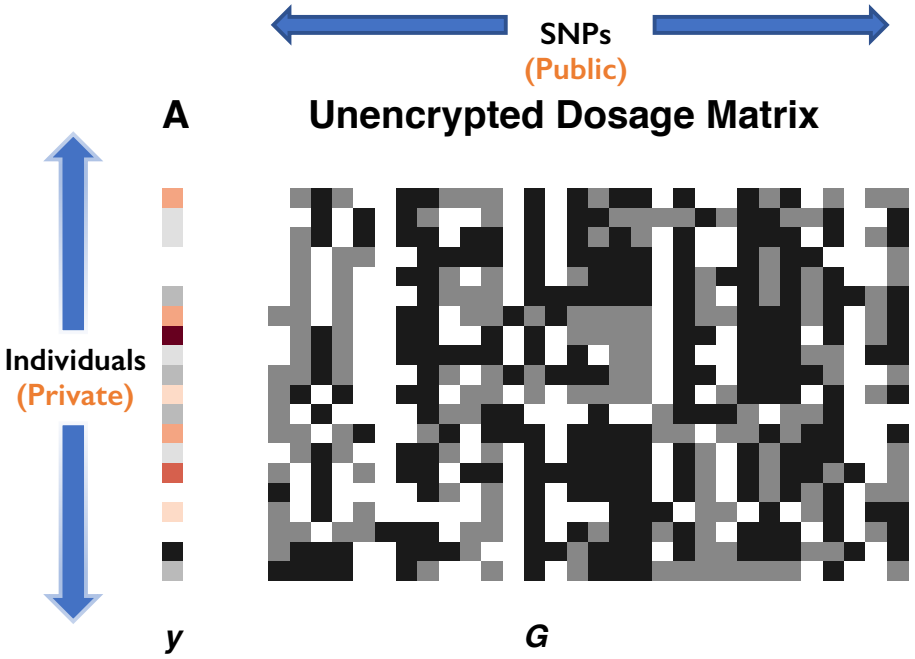
[1] Genetics Institute, University College London, Gower St London WC1E 6BT.

[2]Genetics, Genomics and Informatics, University of Tennessee Health Science Center, 71 S Manassas St, Memphis TN, USA

[3] Department of Statistics, University of Oxford, 29 St Giles', Oxford OX1 3LB, UK

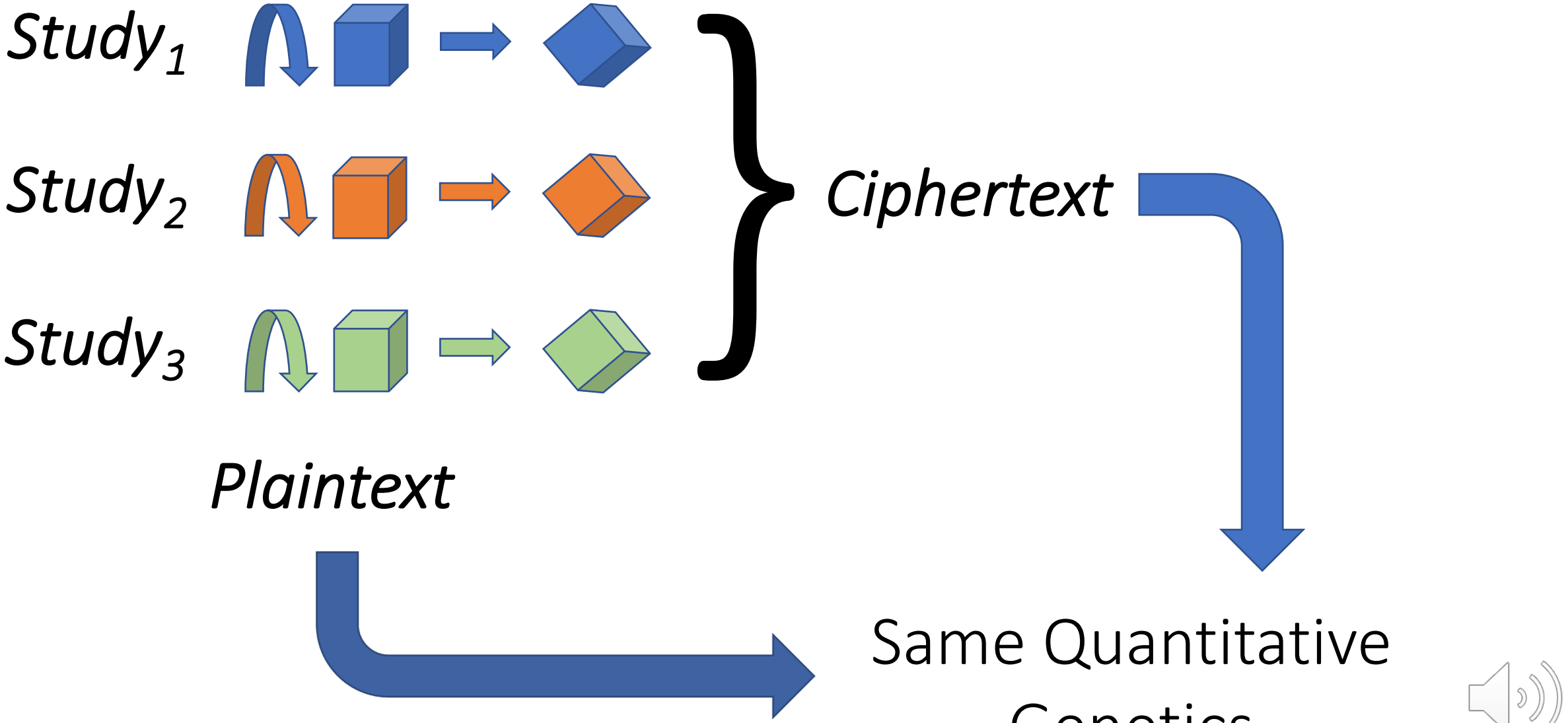
*Corresponding author: r.mott@ucl.ac.uk

Aims



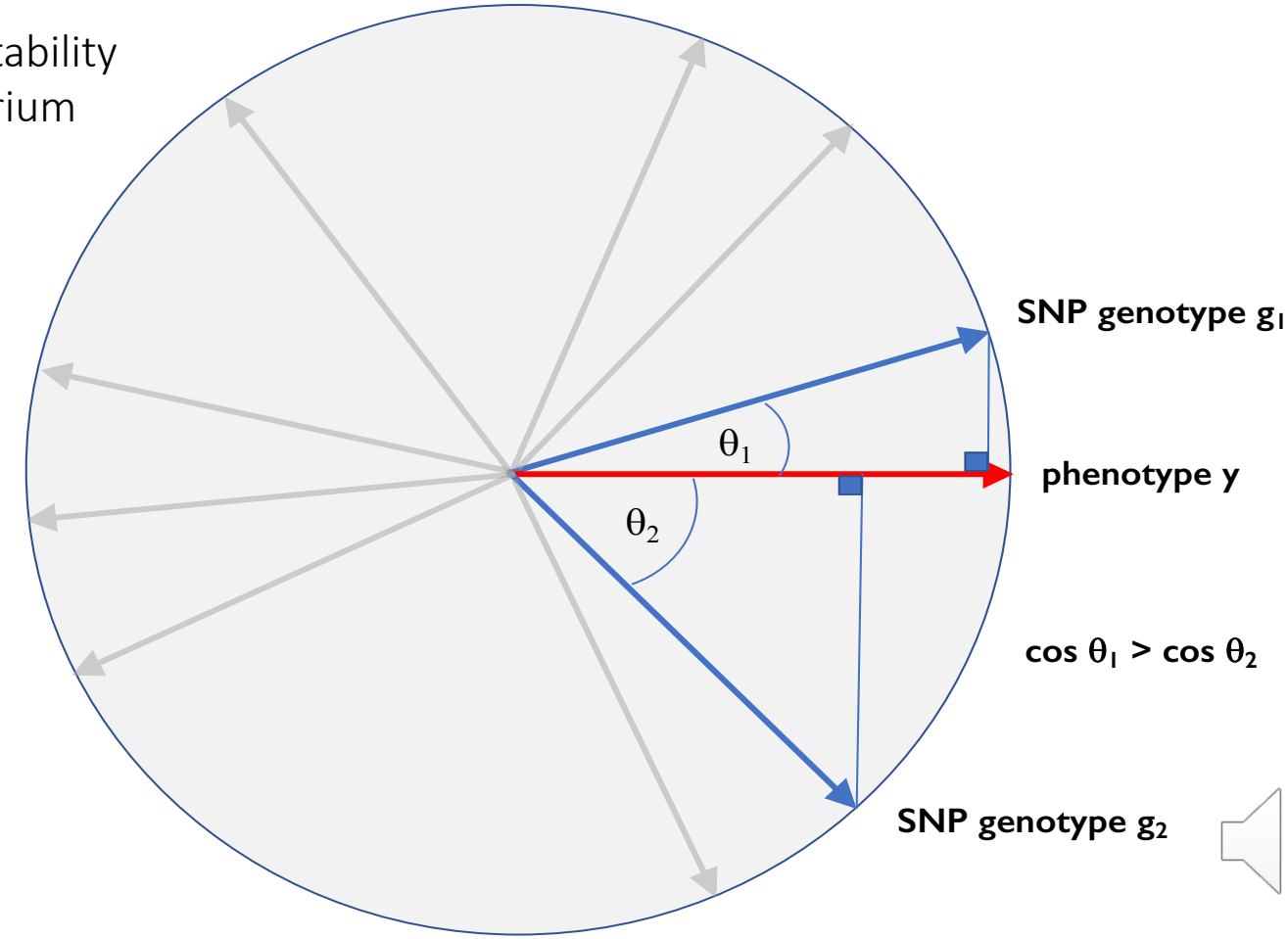
- Disguise genotypes of individuals (rows)
- Preserve relationships between columns:
  - phenotype *vs* genotypes – association, heritability
  - genotypes *vs* genotypes - linkage disequilibrium

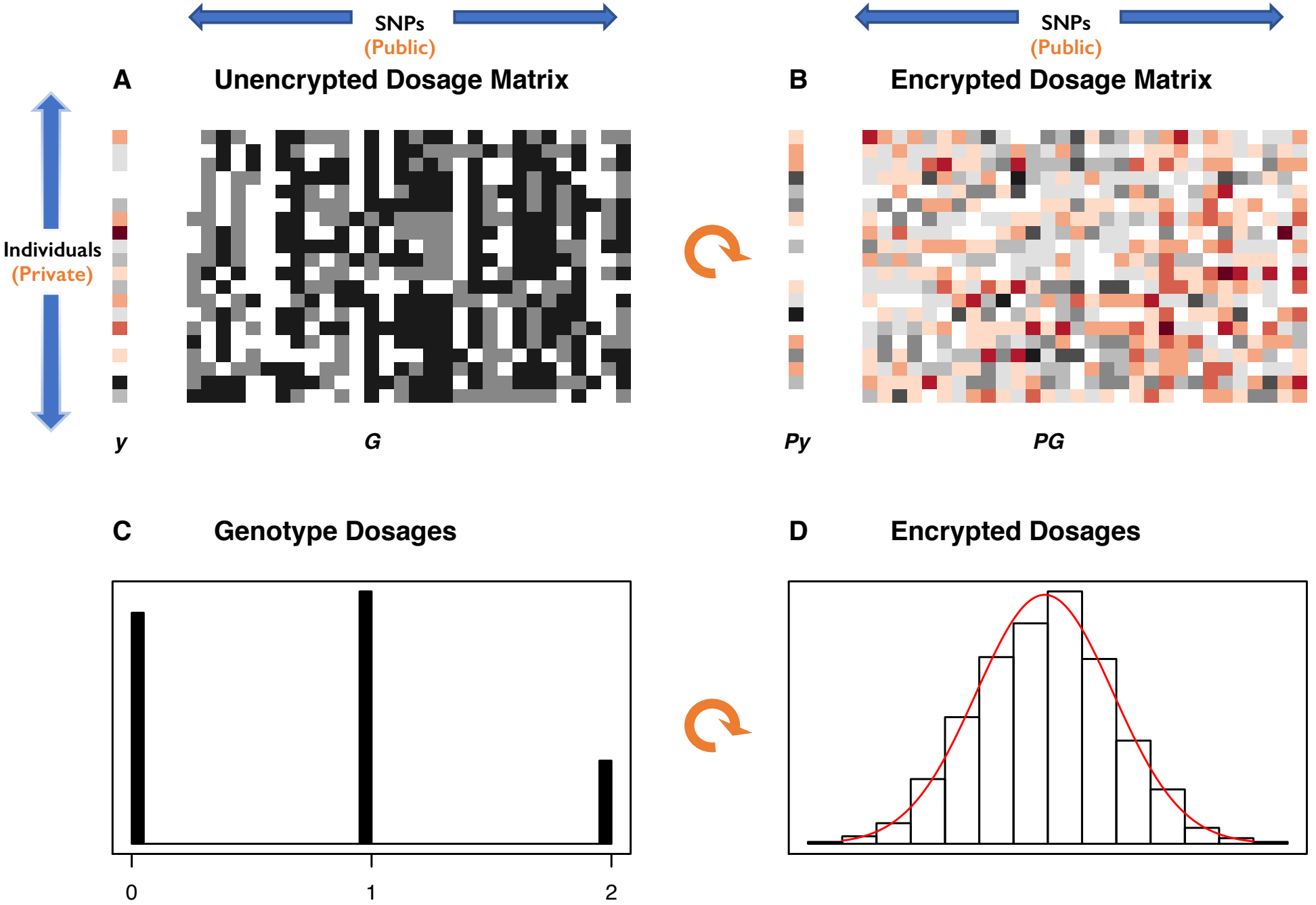
Homomorphic Genotype Encryption

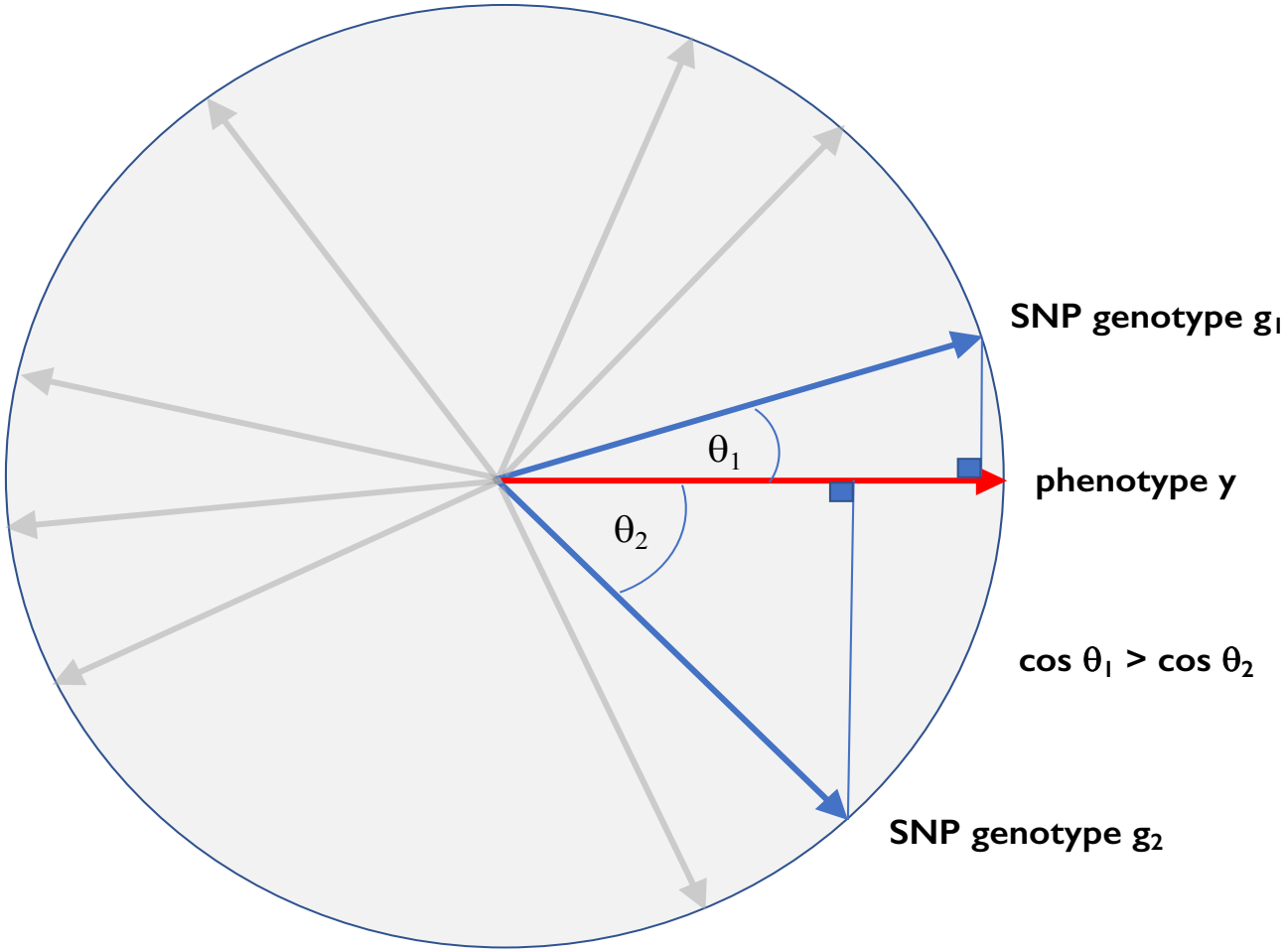


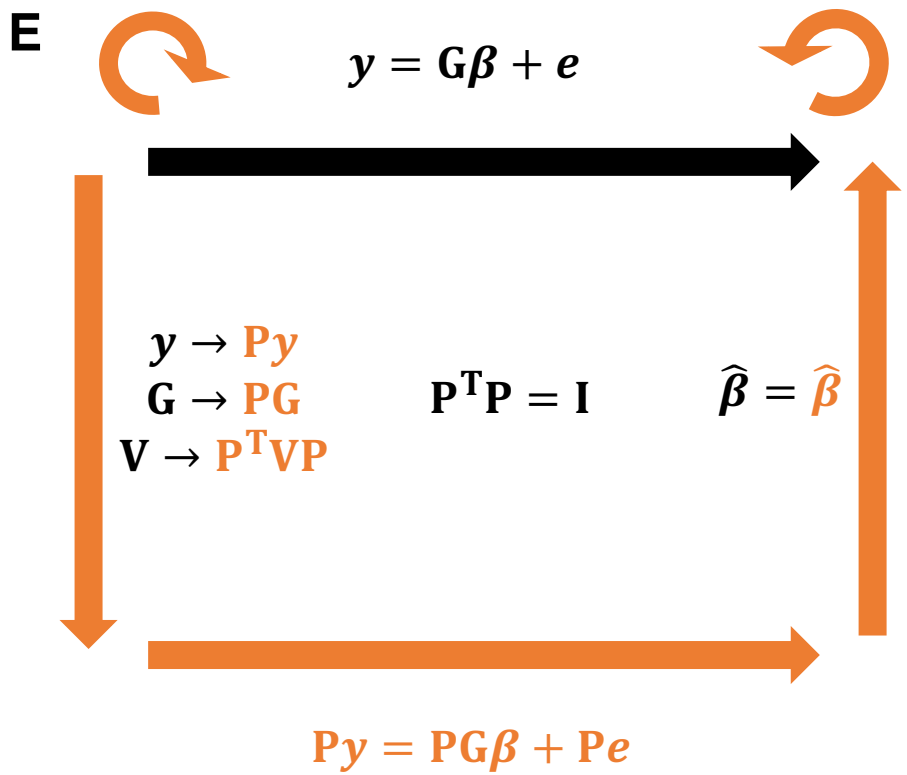
Homomorphic Genotype Encryption

- Disguise genotypes of individuals (rows)
- Preserve relationships between columns:
  - phenotype *vs* genotypes – association, heritability
  - genotypes *vs* genotypes - linkage disequilibrium



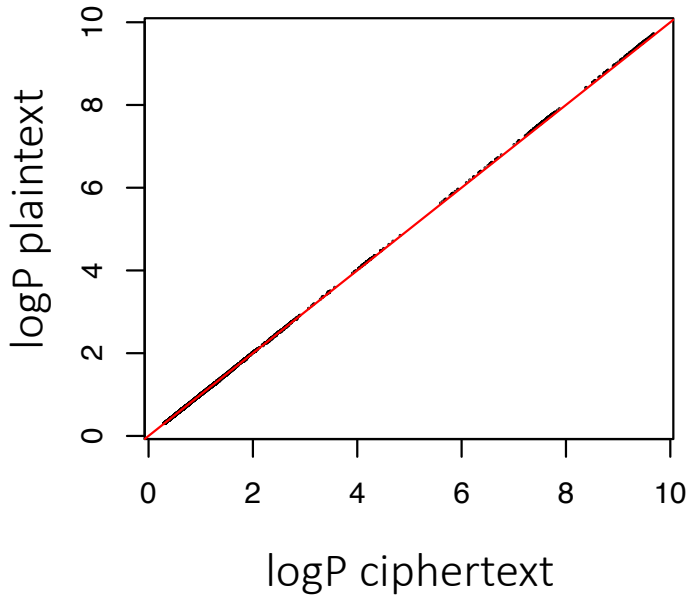

SNPs
(Public)
A
Unencrypted Dosage Matrix
Individuals
(Private)
y
G
SNPs
(Public)
B
Encrypted Dosage Matrix
Py
PG
C
Genotype Dosages
0
1
2
D
Encrypted Dosages


SNP genotype g_1
phenotype y
θ_1
θ_2
$\cos\theta_1 > \cos\theta_2$
SNP genotype g_2


E
$y = G\beta + e$
$y \rightarrow Py$
$G \rightarrow PG$
$V \rightarrow P^TVP$
$P^TP = I$
$\hat{\beta} = \hat{\beta}$
$Py = PG\beta + Pe$

GWAS p-values are unchanged

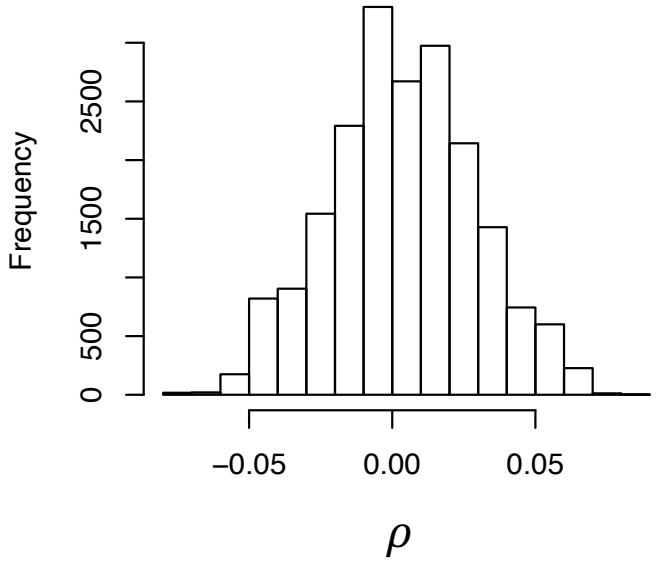
GWAS for Platelet levels in 2000 outbred mice



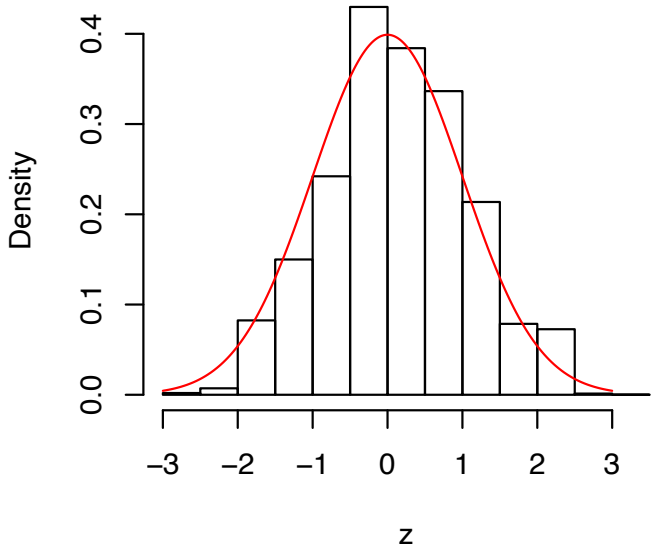
Plaintext $h^2 = 0.0253$
Ciphertext $h^2 = 0.0250$

Random Correlations of plaintext *vs* ciphertext

Pearson Correlations of plaintext and ciphertext

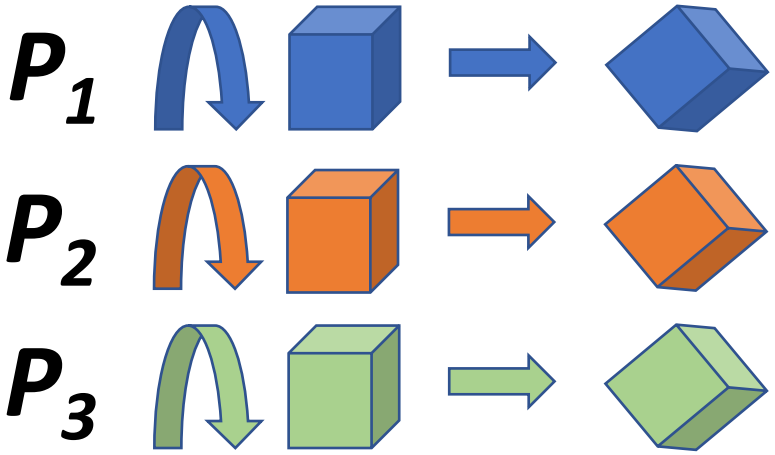


Z-transformed Pearson Correlations

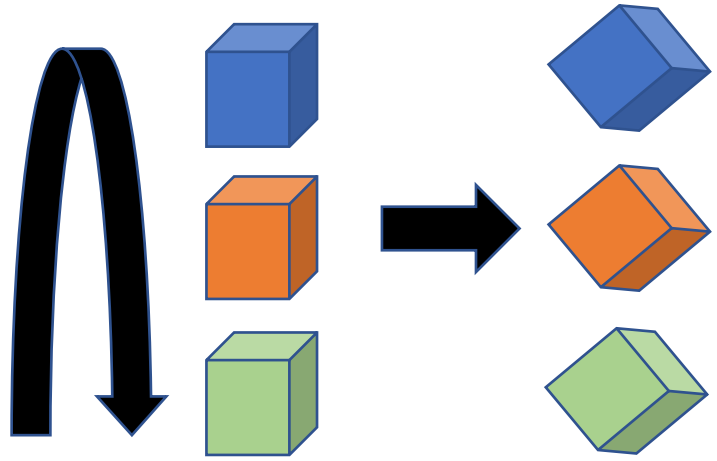


$$z = \rho\sqrt{\frac{n-2}{1-\rho^2}} \sim N(0,1)$$

Federated Mega Analysis



$$\begin{pmatrix} P_1 & 0 & 0 \\ 0 & P_2 & 0 \\ 0 & 0 & P_3 \end{pmatrix}$$



Summary

- Orthogonal Transformations map plaintext genotypes to cyphertext
- Preserves Genetic Association estimates, p-values, heritability
- Works for mixed models to control for unequal relatedness
- Its security is presented as a **challenge to the community**
  - Decrypting N individuals requires finding NxN orthogonal matrix key
  - N(N-1)/2 free parameters: (e.g. N=10k implies 50M parameters)
  - We don’t know if decryption is possible, but it is certainly hard.
  - Can you find a way to decrypt?

- BioRXiv https://doi.org/10.1101/2020.04.02.021865
- Paper accepted for publication in Genetics
- https://github.com/encryption4genetics

Funding and Acknowledgements

BBSRC Grants BB/S017372/1, BB/R01356X/1, BB/P024726/1, BB/M011585/1

UT Center for Integrative and Translational Genomics and funds from the UT-ORNL Governor's Chair, NIGMS grant R01GM123489 and NIDA grant P30DA044223.

Thanks to Rob Williams, University of Tennessee