

Single- and cross-tissue heritability of gene expression  
via identity-by-descent in related or unrelated individuals



Alkes L. Price  
Harvard School of Public Health  
November 5, 2010

# Conflict of Interest Disclosure

Four of the authors (Helgason, Thorleifsson, Kong, Stefansson) are shareholders and/or employees of deCODE Genetics, a biotechnology company.



# What is “heritability”?

$$\text{Phenotype (P)} = \text{Genotype (G)} + \text{Environment (E)} \quad (1)$$

The variance of the observable phenotypes ( $\sigma_P^2$ ) can be expressed as a sum of unobserved underlying variances ( $\sigma_G^2$  and  $\sigma_E^2$ ):

$$\sigma_P^2 = \sigma_G^2 + \sigma_E^2 \quad (2)$$

Heritability is defined as a ratio of variances, by expressing the proportion of the phenotypic variance that can be attributed to variance of genotypic values:

$$\text{Heritability (broad sense)} = H^2 = \frac{\sigma_G^2}{\sigma_P^2}$$

The genetic variance can be partitioned into the variance of additive genetic effects (breeding values;  $\sigma_A^2$ ), of dominance (interactions between alleles at the same locus) genetic effects ( $\sigma_D^2$ ), and of epistatic (interactions between alleles at different loci) genetic effects ( $\sigma_I^2$ ):

$$\sigma_G^2 = \sigma_A^2 + \sigma_D^2 + \sigma_I^2$$

$$\text{and heritability (narrow or strict sense)} = h^2 = \frac{\sigma_A^2}{\sigma_P^2}$$



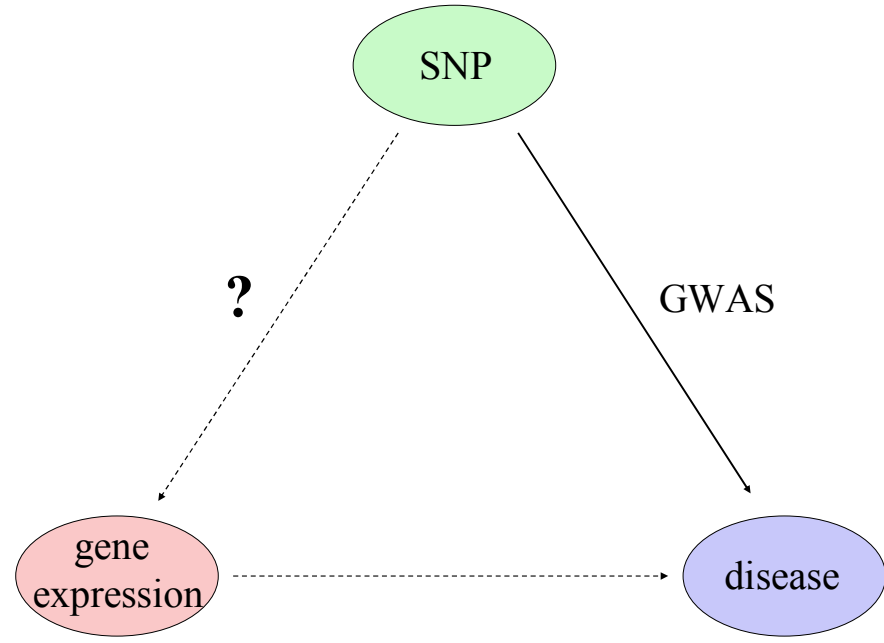
## The case of the missing heritability

When scientists opened up the human genome, they expected to find the genetic components of common traits and diseases. But they were nowhere to be seen. **Brendan Maher** shines a light on six places where the missing loot could be stashed away.

Maher 2008 Nature; Manolio et al. 2009 Nature

# Will studying gene expression help solve the mystery?

- Gene expression mediates the relationship between genotype and disease



- 20,000 gene expression phenotypes may be more statistically informative than 1 (or 5 or 10) disease phenotypes

# Outline

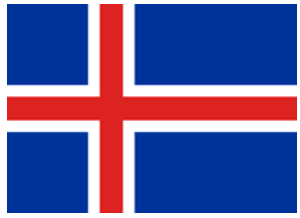
1. **Data sets**
2. Heritability of gene expression in blood and adipose tissue
3. *Cis* and *trans* heritability of gene expression
4. Cross-tissue heritability of gene expression



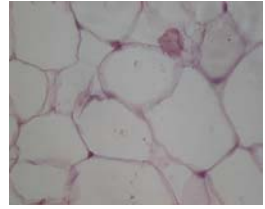
## Icelandic Family Blood (IFB) cohort



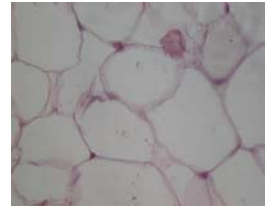
- 687 related Icelandic individuals from family pedigrees
- Blood expression measurements for 18,735 gene transcripts  
(adjusted for gender and age)
- NEW! Illumina 300K genotype data, long-range phased



## Icelandic Family Adipose (IFA) cohort

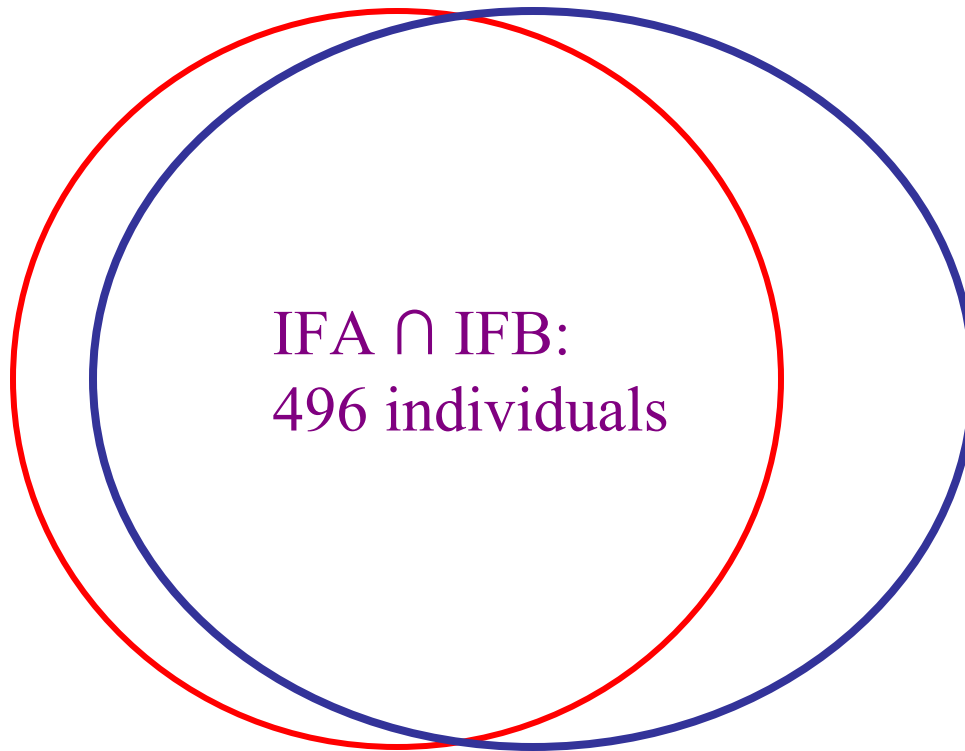


- 531 related Icelandic individuals from family pedigrees
- Adipose tissue expression measurements for 19,099 gene transcripts (adjusted for gender, age and BMI)
- NEW! Illumina 300K genotype data, long-range phased



## IFB and IFA cohorts

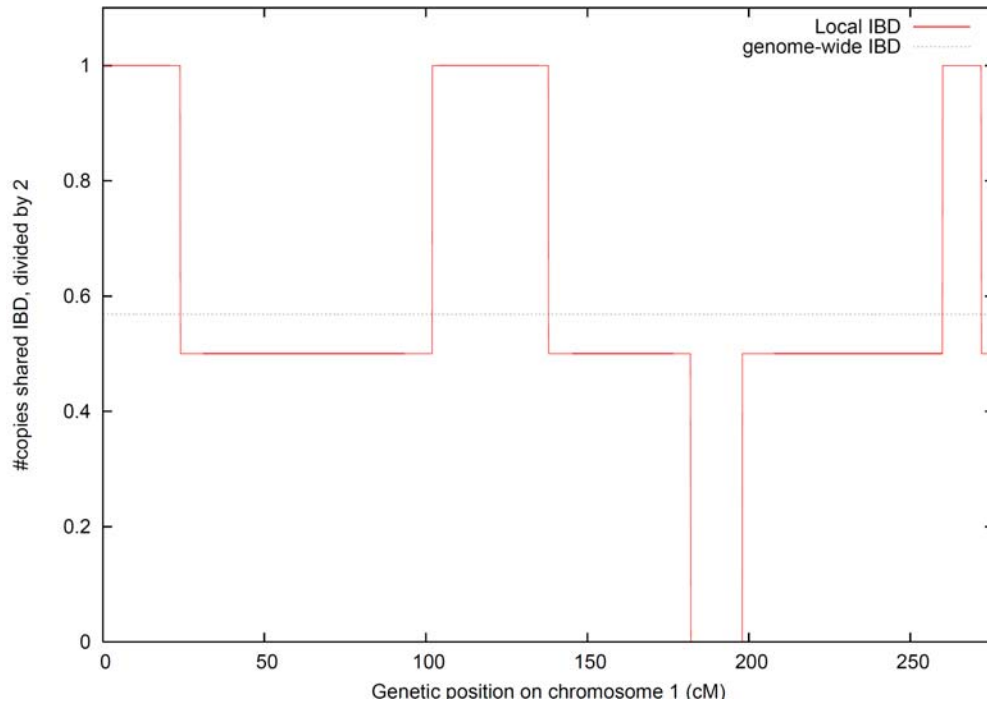
IFA: 531  
individuals  
(adipose)



IFA  $\cap$  IFB:  
496 individuals

IFB: 687  
individuals  
(blood)

# Identity-by-descent (IBD) estimates



actual chr1 local IBD of  
two Icelandic siblings



- Local IBD varies between 0 and 1 across each chromosome
- Genome-wide IBD is equal to 0.568 for this pair of siblings
- Availability of long-range phased data aids IBD estimation
- IBD can be estimated for related or unrelated individuals

Purcell et al. 2007 AJHG; Gusev et al. 2009 Genome Res; Browning et al. 2010 AJHG  
Also see Talk #368 (Browning, 4:45pm Fri Nov 5), Talk #369 (Gusev, 5pm Fri Nov 5)

# Outline

1. Data sets
- 2. Heritability of gene expression in blood and adipose tissue**
3. *Cis* and *trans* heritability of gene expression
4. Cross-tissue heritability of gene expression

# Heritability of gene expression in IFB, IFA

## VARIANCE-COMPONENTS MODEL:

$\Theta$  = matrix of genome-wide IBD for each pair of individuals

$e_{gs}$  = normalized gene expression for gene  $g$ , individual  $s$

$V_g$  = covariance matrix of normalized gene expression for gene  $g$

$$V_g = h_g^2 \Theta + (1 - h_g^2) I$$

We fit  $h_g^2$  to maximize  $L(e_{gs} | V_g)$ .

e.g.

$$\Theta = \begin{bmatrix} 1 & 0.5 & 0.5 \\ 0.5 & 1 & 0.5 \\ 0.5 & 0.5 & 1 \end{bmatrix}, h_g^2 = 0.20 \Rightarrow V_g = \begin{bmatrix} 1 & 0.1 & 0.1 \\ 0.1 & 1 & 0.1 \\ 0.1 & 0.1 & 1 \end{bmatrix}$$

# Heritability of gene expression in IFB, IFA

## VARIANCE-COMPONENTS MODEL:

$\Theta$  = matrix of genome-wide IBD for each pair of individuals

$e_{gs}$  = normalized gene expression for gene  $g$ , individual  $s$

$V_g$  = covariance matrix of normalized gene expression for gene  $g$

$$V_g = h_g^2 \Theta + (1 - h_g^2) I$$

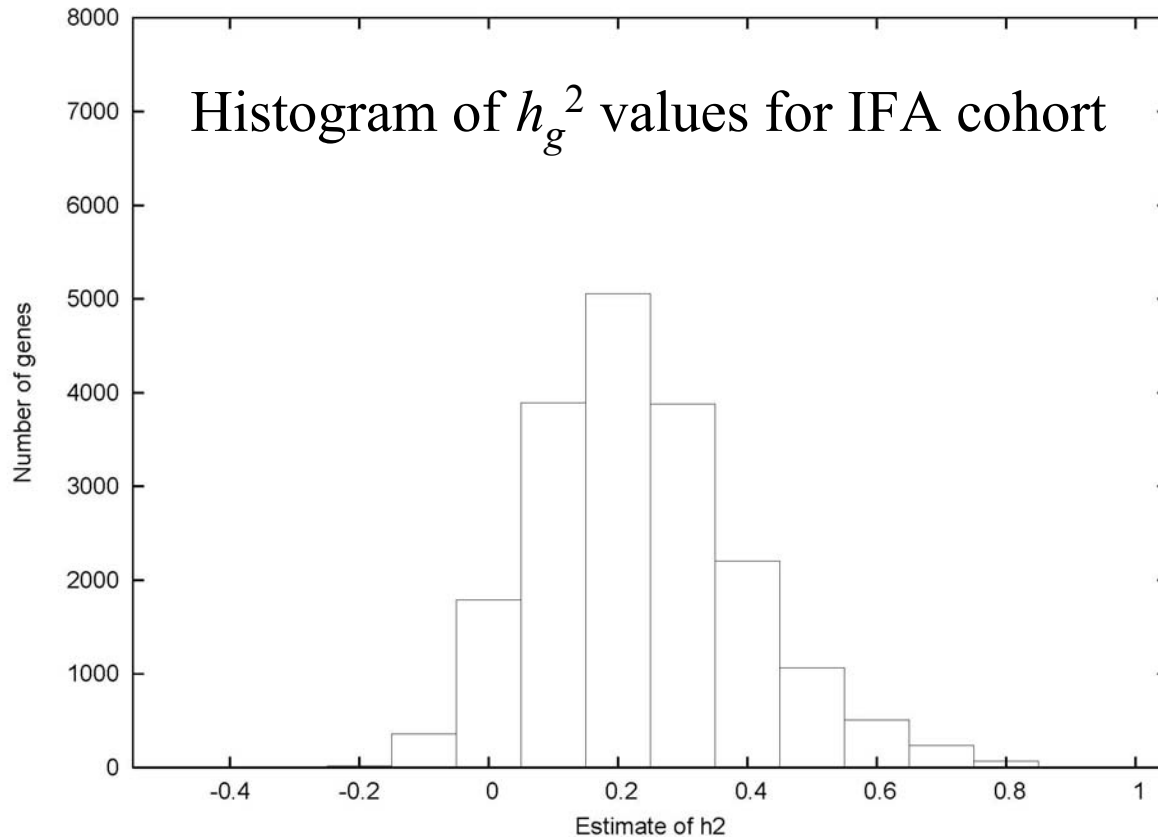
We fit  $h_g^2$  to maximize  $L(e_{gs} | V_g)$ .

## THE RESULTS:

Average  $h^2 = 0.150 \pm 0.011$  for IFB (blood)

Average  $h^2 = 0.234 \pm 0.011$  for IFA (adipose tissue)

# Significant excess of genes with $h_g^2 > 0$



IFB: 42% of genes had  $h_g^2$  significantly ( $P < 0.05$ ) greater than 0  
IFA: 63% of genes had  $h_g^2$  significantly ( $P < 0.05$ ) greater than 0

$h^2 = 0.15-0.23$  is consistent with previous results

- Heritability  $> 0.20$  for 10-13% of LCL gene expression transcripts  
(Stranger et al. 2007 Nat Genet)
- Mean heritability of 0.203 across LCL gene expression transcripts  
(Dixon et al. 2007 Nat Genet)
- Heritability  $> 0$  for 60-70% of IFB, IFA gene expression transcripts  
Average heritability of 0.30 for transcripts with heritability  $> 0$   
(Emilsson et al. 2008 Nature)

# Outline

1. Data sets
2. Heritability of gene expression in blood and adipose tissue
3. ***Cis* and *trans* heritability of gene expression**
4. Cross-tissue heritability of gene expression

# *cis* vs. *trans* heritability in IFB, IFA

## VARIANCE-COMPONENTS MODEL:

$\Theta$  = matrix of genome-wide IBD for each pair of individuals

$\Gamma_g$  = matrix of local IBD at gene  $g$  for each pair of individuals

$e_{gs}$  = normalized gene expression for gene  $g$ , individual  $s$

$V_g$  = covariance matrix of normalized gene expression for gene  $g$

$$V_g = h_{g,cis}^2 \Gamma_g + h_{g,trans}^2 \Theta + (1 - h_{g,cis}^2 - h_{g,trans}^2) I$$

We fit  $h_{g,cis}^2$  and  $h_{g,trans}^2$  to maximize  $L(e_{gs} | V_g)$ .

# *cis* vs. *trans* heritability in IFB, IFA

## VARIANCE-COMPONENTS MODEL:

$\Theta$  = matrix of genome-wide IBD for each pair of individuals

$\Gamma_g$  = matrix of local IBD at gene  $g$  for each pair of individuals

$e_{gs}$  = normalized gene expression for gene  $g$ , individual  $s$

$V_g$  = covariance matrix of normalized gene expression for gene  $g$

$$V_g = h_{g,cis}^2 \Gamma_g + h_{g,trans}^2 \Theta + (1 - h_{g,cis}^2 - h_{g,trans}^2) I$$

We fit  $h_{g,cis}^2$  and  $h_{g,trans}^2$  to maximize  $L(e_{gs} | V_g)$ .

## THE RESULTS:

Average  $h_{cis}^2 = 0.055 \pm 0.001$ ,  $h_{trans}^2 = 0.095 \pm 0.010$  for IFB (blood)

Average  $h_{cis}^2 = 0.057 \pm 0.002$ ,  $h_{trans}^2 = 0.177 \pm 0.010$  for IFA (adipose)

# *cis* vs. *trans* heritability in IFB, IFA

## VARIANCE-COMPONENTS MODEL:

$\Theta$  = matrix of genome-wide IBD for each pair of individuals

$\Gamma_g$  = matrix of local IBD at gene  $g$  for each pair of individuals

$e_{gs}$  = normalized gene expression for gene  $g$ , individual  $s$

$V_g$  = covariance matrix of normalized gene expression for gene  $g$

$$V_g = h_{g,cis}^2 \Gamma_g + h_{g,trans}^2 \Theta + (1 - h_{g,cis}^2 - h_{g,trans}^2) I$$

We fit  $h_{g,cis}^2$  and  $h_{g,trans}^2$  to maximize  $L(e_{gs} | V_g)$ .

## THE RESULTS:

Average  $h_{cis}^2 = 0.055 \pm 0.001$ ,  $h_{trans}^2 = 0.095 \pm 0.010$  for IFB (blood)

Average  $h_{cis}^2 = 0.057 \pm 0.002$ ,  $h_{trans}^2 = 0.177 \pm 0.010$  for IFA (adipose)

# *cis* vs. *trans* heritability in IFB, IFA

## VARIANCE-COMPONENTS MODEL:

$\Theta$  = matrix of genome-wide IBD for each pair of individuals

$\Gamma_g$  = matrix of local IBD at gene  $g$  for each pair of individuals

$e_{gs}$  = normalized gene expression for gene  $g$ , individual  $s$

$V_g$  = covariance matrix of normalized gene expression for gene  $g$

$$V_g = h_{g,cis}^2 \Gamma_g + h_{g,trans}^2 \Theta + (1 - h_{g,cis}^2 - h_{g,trans}^2) I$$

We fit  $h_{g,cis}^2$  and  $h_{g,trans}^2$  to maximize  $L(e_{gs} | V_g)$ .

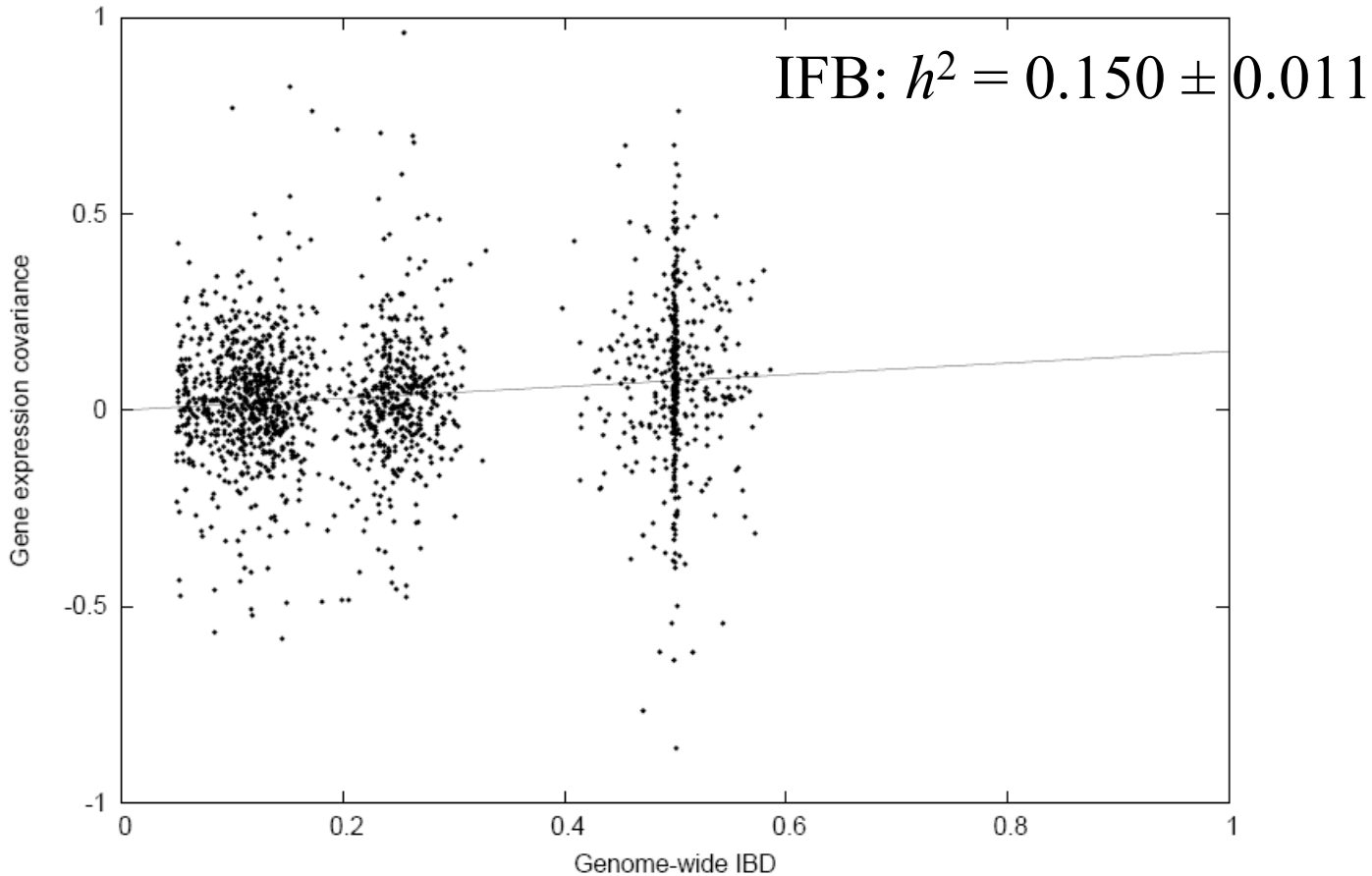
## THE RESULTS:

Average  $h_{cis}^2 = 0.055 \pm 0.001$ ,  $h_{trans}^2 = 0.095 \pm 0.010$  for IFB (blood)

Average  $h_{cis}^2 = 0.057 \pm 0.002$ ,  $h_{trans}^2 = 0.177 \pm 0.010$  for IFA (adipose)

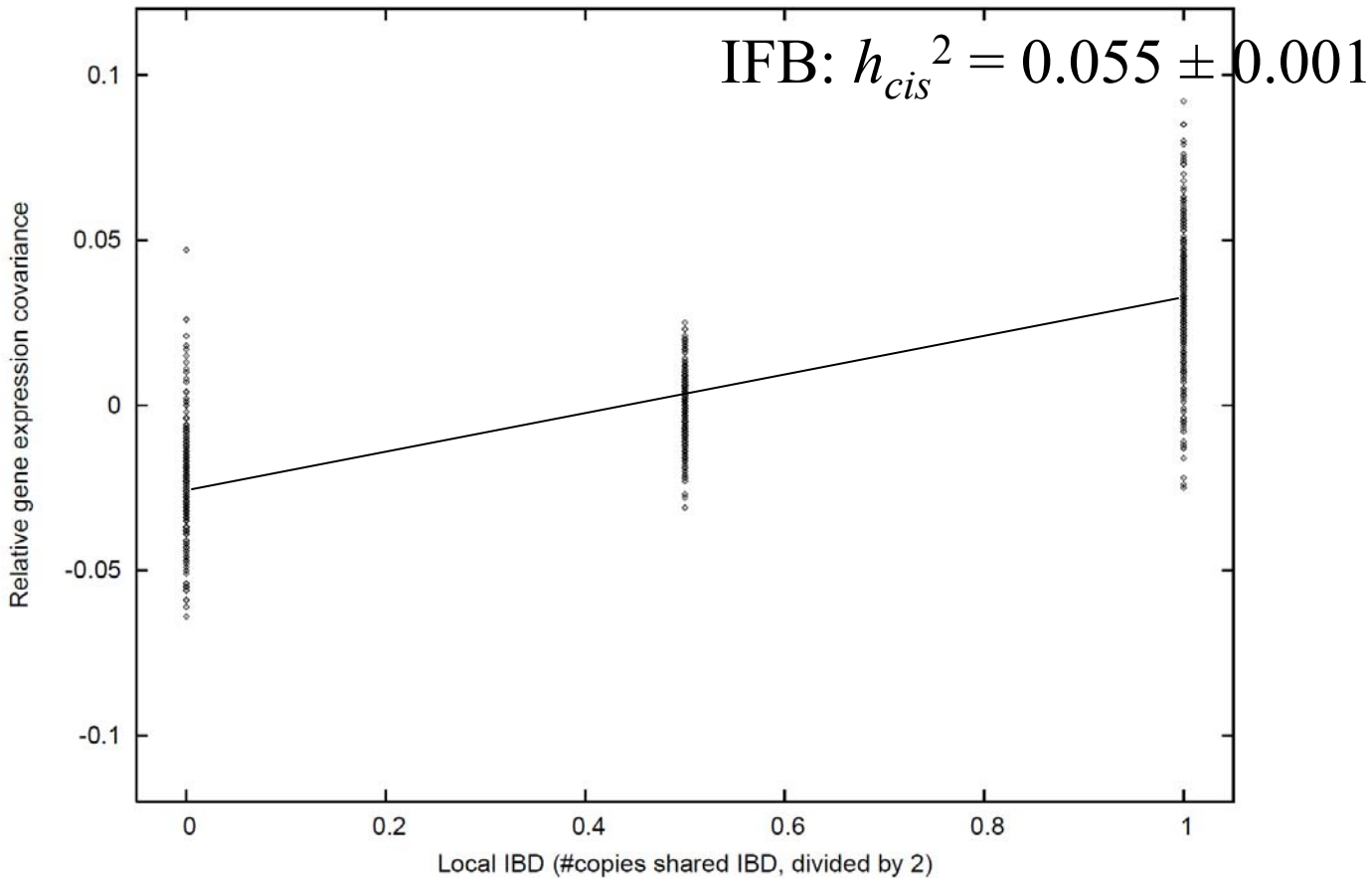
# Systematic noise covariance hinders estimation of $h^2$

$\text{Cov}_g(e_{gs}, e_{gt})$  vs.  $\Theta$  (each point is one pair of individuals  $s, t$ )



# No systematic noise covariance in estimation of $h_{cis}^2$

Relative  $\text{Cov}_g(e_{gs}, e_{gt})$  vs.  $\Gamma_g$  (each point is one pair of siblings  $s, t$ )



# *cis* vs. *trans* heritability in IFB, IFA

## VARIANCE-COMPONENTS MODEL:

$\Theta$  = matrix of genome-wide IBD for each pair of individuals

$\Gamma_g$  = matrix of local IBD at gene  $g$  for each pair of individuals

$e_{gs}$  = normalized gene expression for gene  $g$ , individual  $s$

$V_g$  = covariance matrix of normalized gene expression for gene  $g$

$$V_g = h_{g,cis}^2 \Gamma_g + h_{g,trans}^2 \Theta + (1 - h_{g,cis}^2 - h_{g,trans}^2) I$$

We fit  $h_{g,cis}^2$  and  $h_{g,trans}^2$  to maximize  $L(e_{gs} | V_g)$ .

## THE RESULTS:

$$\begin{array}{l} \text{Average } h_{cis}^2 = 0.055, h_{trans}^2 = 0.095 \Rightarrow \pi_{cis} = 37\% \text{ for IFB (blood)} \\ \text{Average } h_{cis}^2 = 0.057, h_{trans}^2 = 0.177 \Rightarrow \pi_{cis} = 24\% \text{ for IFA (adipose)} \end{array}$$

Is  $\pi_{cis} = 24-37\%$  consistent with previous results?

- Cheung et al. 2010 PLoS Biol, experimental validation of *cis*- and *trans*-acting regulators: “The majority of regulators act in *trans*”.

## Is $\pi_{cis} = 24-37\%$ consistent with previous results?

- Cheung et al. 2010 PLoS Biol, experimental validation of *cis*- and *trans*-acting regulators: “The majority of regulators act in *trans*”.
- Admixture analysis:  $12 \pm 3\%$  of heritable variation in LCL gene expression is due to *cis* variants (Price et al. 2008 PLoS Genet)
- Admixture analysis of LCL gene expression in HapMap 3 MEX (HapMap 3 Consortium 2010 Nature; Stranger et al. submitted): Similar results:  $13 \pm 9\%$  of heritable variation is due to *cis* variants.

## Is $\pi_{cis} = 24-37\%$ consistent with previous results?

- Cheung et al. 2010 PLoS Biol, experimental validation of *cis*- and *trans*-acting regulators: “The majority of regulators act in *trans*”.
- Admixture analysis:  $12 \pm 3\%$  of heritable variation in LCL gene expression is due to *cis* variants (Price et al. 2008 PLoS Genet)
- Admixture analysis of LCL gene expression in HapMap 3 MEX (HapMap 3 Consortium 2010 Nature; Stranger et al. submitted): Similar results:  $13 \pm 9\%$  of heritable variation is due to *cis* variants.

Tissue-dependence?

GxG interactions?

Epigenetic effects?

## Is $\pi_{cis} = 24-37\%$ consistent with previous results?

- Cheung et al. 2010 PLoS Biol, experimental validation of *cis*- and *trans*-acting regulators: “The majority of regulators act in *trans*”.
- Admixture analysis:  $12 \pm 3\%$  of heritable variation in LCL gene expression is due to cis variants (Price et al. 2008 PLoS Genet)
- Admixture analysis of LCL gene expression in HapMap 3 MEX (HapMap 3 Consortium 2010 Nature; Stranger et al. submitted): Similar results:  $13 \pm 9\%$  of heritable variation is due to cis variants.

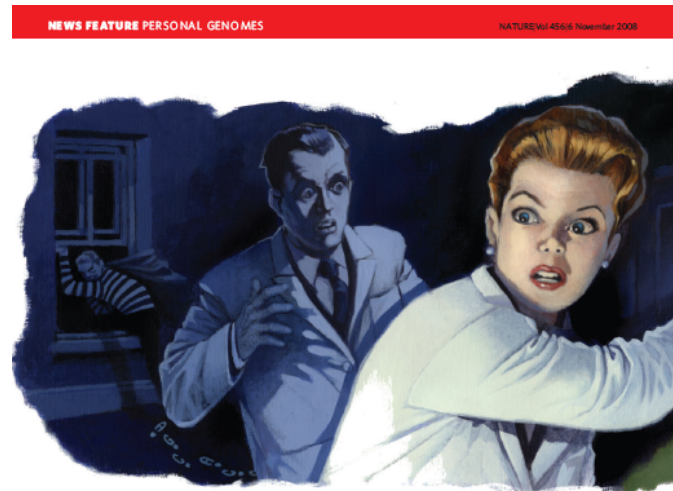
Tissue-dependence?

GxG interactions?

Epigenetic effects?

# The epigenetic hypothesis

“Epigenetics, changes in gene expression that are inherited but not caused by changes in genetic sequences, confuses things further ... the idea that grandma’s environment could affect future generations is controversial—and such effects would have been included in the heritability normally attributed to genes.”



NEWS FEATURE PERSONAL GENOMES

NATURE 456 6 November 2008

## The case of the missing heritability

When scientists opened up the human genome, they expected to find the genetic components of common traits and diseases. But they were nowhere to be seen. **Brendan Maher** shines a light on six places where the missing loot could be stashed away.

# The epigenetic hypothesis

---

## DNA methylation profiles in monozygotic and dizygotic twins

Zachary A Kaminsky<sup>1,2</sup>, Thomas Tang<sup>1</sup>, Sun-Chong Wang<sup>1,3</sup>, Carolyn Ptak<sup>1,2</sup>, Gabriel H T Oh<sup>1,2</sup>, Albert H C Wong<sup>1,2</sup>, Laura A Feldcamp<sup>1,2</sup>, Carl Virtanen<sup>4</sup>, Jonas Halfvarson<sup>5,6</sup>, Curt Tysk<sup>5,6</sup>, Allan F McRae<sup>7</sup>, Peter M Visscher<sup>7</sup>, Grant W Montgomery<sup>7</sup>, Irving I Gottesman<sup>8</sup>, Nicholas G Martin<sup>7</sup> & Art Petronis<sup>1,2</sup>

“Molecular mechanisms of heritability may not be limited to DNA sequence differences.”

# Similar $h_{cis}^2$ using IBD in unrelated individuals

## VARIANCE-COMPONENTS MODEL:

$\Theta$  = matrix of genome-wide IBD for each pair of individuals

$\Gamma_g$  = matrix of local IBD at gene  $g$  for each pair of individuals

$e_{gs}$  = normalized gene expression for gene  $g$ , individual  $s$

$V_g$  = covariance matrix of normalized gene expression for gene  $g$

$$V_g = h_{g,cis}^2 \Gamma_g + h_{g,trans}^2 \Theta + (1 - h_{g,cis}^2 - h_{g,trans}^2) I$$

We fit  $h_{g,cis}^2$  and  $h_{g,trans}^2$  to maximize  $L(e_{gs} | V_g)$ .

## APPLICATION TO SUBSETS OF UNRELATED INDIVIDUALS:

Average  $h_{cis}^2 = 0.057$  for IFB (blood)

Average  $h_{cis}^2 = 0.067$  for IFA (adipose tissue)

## Is $\pi_{cis} = 24-37\%$ consistent with previous results?

- Cheung et al. 2010 PLoS Biol, experimental validation of *cis*- and *trans*-acting regulators: “The majority of regulators act in *trans*”.
- Admixture analysis:  $12 \pm 3\%$  of heritable variation in LCL gene expression is due to *cis* variants (Price et al. 2008 PLoS Genet)
- Admixture analysis of LCL gene expression in HapMap 3 MEX (HapMap 3 Consortium 2010 Nature; Stranger et al. submitted): Similar results:  $13 \pm 9\%$  of heritable variation is due to *cis* variants.

Tissue-dependence?

GxG interactions?

Transgenerational epigenetic inheritance does not play a major role in heritability of gene expression.

see Youngson & Whitelaw 2008  
Annu Rev Genomics Hum Genet

# Outline

1. Data sets
2. Heritability of gene expression in blood and adipose tissue
3. *Cis* and *trans* heritability of gene expression
4. **Cross-tissue heritability of gene expression**



## Gene expression is correlated across tissues



Let  $e_{bgs}$  denote blood gene expression for gene  $g$ , individual  $s$

Let  $e_{ags}$  denote adipose gene expression for gene  $g$ , individual  $s$

$$\rho(e_{bgs}, e_{ags}) = 0.041 \quad (\text{P-value} \ll 10^{-12})$$

Why? Is this due to heritable factors?

# Cross-tissue heritability in IFB, IFA

THE BASIC IDEA:

Single-tissue heritability:

Blood expression of individual  $s$   $\Leftrightarrow$  Blood expression of relative of  $s$

Cross-tissue heritability:

Blood expression of individual  $s$   $\Leftrightarrow$  **Adipose** expression of relative of  $s$

e.g. cross-heritability of  
two correlated height phenotypes:  
Macgregor et al. 2006 Hum Genet

# Cross-tissue heritability in IFB, IFA

VARIANCE-COMPONENTS MODEL:

$\Theta$  = matrix of genome-wide IBD,  $\Gamma_g$  = matrix of local IBD at gene  $g$

$e_{bgs}$  = normalized gene expression in blood for gene  $g$ , individual  $s$

$e_{ags}$  = normalized gene expression in adipose for gene  $g$ , individual  $s$

$W_g = 2N \times 2N$  covariance matrix of  $2N$ -long vector  $(e_{bg}, e_{ag})$  for gene  $g$

$$X_{g,cis} = \begin{pmatrix} h_{bg,cis}^2 & \xi_{g,cis}^2 \\ \xi_{g,cis}^2 & h_{ag,cis}^2 \end{pmatrix}, X_{g,trans} = \begin{pmatrix} h_{bg,trans}^2 & \xi_{g,trans}^2 \\ \xi_{g,trans}^2 & h_{ag,trans}^2 \end{pmatrix},$$

$$X_{g,env} = \begin{pmatrix} 1 - h_{bg,cis}^2 - h_{bg,trans}^2 & \rho_g - \xi_{g,cis}^2 - \xi_{g,trans}^2 \\ \rho_g - \xi_{g,cis}^2 - \xi_{g,trans}^2 & 1 - h_{ag,cis}^2 - h_{ag,trans}^2 \end{pmatrix},$$

$$W_g = X_{g,cis} \otimes \Gamma_g + X_{g,trans} \otimes \Theta + X_{g,env} \otimes I,$$

Fit  $h_{bg,cis}^2, h_{bg,trans}^2, h_{ag,cis}^2, h_{ag,trans}^2, \rho_g, \xi_{g,cis}^2, \xi_{g,trans}^2$  to maximize  $L(e_{bgs}, e_{ags} | W_g)$

# Cross-tissue heritability in IFB, IFA

RESULTS:

$$\xi^2 = 0.030 \pm 0.006$$

$$\xi_{cis}^2 = \underbrace{0.031 \pm 0.001,}$$

explains bulk of  $\rho(e_{bgs}, e_{ags})$

$$\xi_{trans}^2 = \underbrace{-0.001 \pm 0.006}$$

not statistically significant

**$\Rightarrow$  Most cross-tissue similarity is heritable and *cis*-regulated**

# Cross-tissue heritability in IFB, IFA

RESULTS:

$$\xi^2 = 0.030 \pm 0.006$$

$$\xi_{cis}^2 = \underbrace{0.031 \pm 0.001,}_{\text{explains bulk of } \rho(e_{bgs}, e_{ags})}$$

$$\xi_{trans}^2 = \underbrace{-0.001 \pm 0.006}_{\text{not statistically significant}}$$

⇒ **Most cross-tissue similarity is heritable and *cis*-regulated**

⇒ **About half of *cis*-regulation is shared across tissues ( $\xi_{cis}^2 / h_{cis}^2$ )**

## Tissue-specific *cis*-regulation is consistent with previous studies

- >50% of *cis* linkage signals overlap between blood and adipose (Emilsson et al. 2008 Nature)
- 54% / 50% / 54% of *cis* eQTLs in fibroblasts / LCLs / T cells are cell-type specific (Dimas et al. 2009 Science)
- 11-22% of genes show allele-specific expression in four cell lines, of which 4.3-8.5% were tissue-specific (Zhang et al. 2009 Nat Methods)

## Is $\pi_{cis} = 24-37\%$ consistent with previous results?

- Cheung et al. 2010 PLoS Biol, experimental validation of *cis*- and *trans*-acting regulators: “The majority of regulators act in *trans*”.
- Admixture analysis:  $12 \pm 3\%$  of heritable variation in LCL gene expression is due to cis variants (Price et al. 2008 PLoS Genet)
- Admixture analysis of LCL gene expression in HapMap 3 MEX (HapMap 3 Consortium 2010 Nature; Stranger et al. submitted): Similar results:  $13 \pm 9\%$  of heritable variation is due to cis variants.

**Tissue-dependence?**

GxG interactions?

Cell-type specific *cis*-regulation could explain tissue-dependent  $\pi_{cis}$

Admixture  $\Rightarrow$  12% *cis* (**LCLs**)

Family heritability  $\Rightarrow$  24-37% *cis* (**blood + adipose tissue**)

LCLs

1 cell type

*cis* effects

+  
*trans* effects

blood + adipose tissue

many cell types (heterogeneous)

shared *cis* effects  $\Rightarrow$  similar  $h_{cis}^2$

+  
no shared *trans* effects  $\Rightarrow \downarrow h_{trans}^2$

$\Rightarrow \uparrow \pi_{cis}^2$

# Conclusions

- Gene expression is heritable: average  $h^2 = 0.15, 0.23$  (blood, adipose)

# Conclusions

- Gene expression is heritable: average  $h^2 = 0.15, 0.23$  (blood, adipose)
- A fraction of heritability is due to *cis*-regulation:  $\pi_{cis}^2 = 37\%, 24\%$

# Conclusions

- Gene expression is heritable: average  $h^2 = 0.15, 0.23$  (blood, adipose)
- A fraction of heritability is due to *cis*-regulation:  $\pi_{cis}^2 = 37\%, 24\%$
- Most cross-tissue similarity is heritable and *cis*-regulated  
(about half of *cis*-regulation is shared across blood and adipose tissue)

# Conclusions

- Gene expression is heritable: average  $h^2 = 0.15, 0.23$  (blood, adipose)
- A fraction of heritability is due to *cis*-regulation:  $\pi_{cis}^2 = 37\%, 24\%$
- Most cross-tissue similarity is heritable and *cis*-regulated  
(about half of *cis*-regulation is shared across blood and adipose tissue)
- Cell-type specific *cis*-regulation could explain higher  $\pi_{cis}^2$  in more heterogeneous tissue types

# Acknowledgements

HSPH:  
Alkes Price



HMS:  
Steve McCarroll

deCODE Genetics:  
Agnar Helgason  
Gudmar Thorleifsson  
Augustine Kong  
Kari Stefansson

... with additional thanks to David Reich, Aditi Hazra, Barbara Stranger, Manolis Dermitzakis, HapMap 3 Consortium, Benji Raby, Noah Zaitlen, and the Icelandic individuals who participated in the IFB and IFA cohorts.

**Post-doctoral positions available: Google “Alkes HSPH”**