Genomic Imprinting in the Human Brain

Attila Gulyás-Kovács

Chess Lab

Contents

First part
Second part
Imprinting and allelic bias
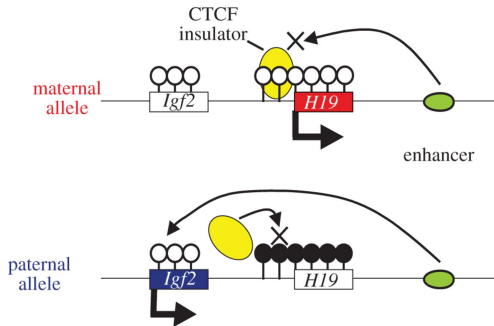Imprinted genes in the brain

Contents

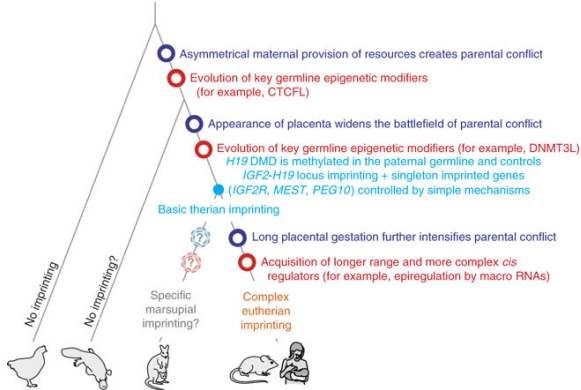# Imprinting and allelic bias

1. epigenetic mechanism
2. variation across age and tissue
3. biological function

Renfree et al 2012 Philos Trans R Soc Lond B
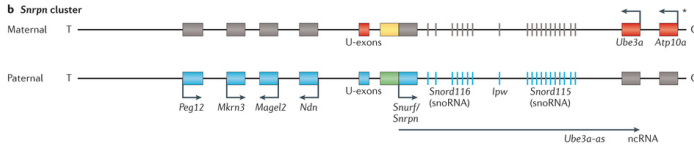
Increasing evolutionary prevalence

Smits et al 2008 Nat Genet

How many imprinted genes?
$\approx 100 \leftrightarrow \approx 1300$ [1]

[1]Gregg et al 2010

Dysfunction: development and growth
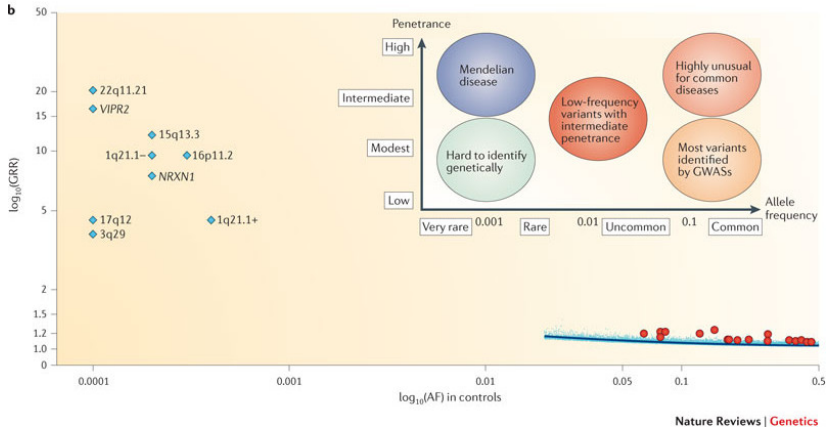

Peters 2014

Angelman syndrome. Boy with a puppet

Prader-Willi syndr. Eugenia "La Monstrua"

Dysfunction: psychology

Angelman/Prader-Willi region implicated in schizophrenia

Nature Reviews | Genetics

Sullivan 2012 Nat Rev Genet.

Questions

1. Imprinted genes in the DLPFC[2]
   - *How many?*
   - *Novelty?*

2. Dependence of allelic bias...
   on age, psychiatric condition, genetics (ancestry), gender
   - *Signal vs noise?*
   - *Genes affected uniformly?*
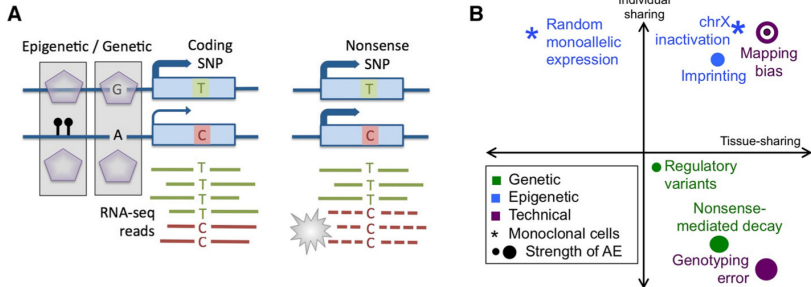   - *Most prominent effects?*

---

[2]dorsolateral prefrontal cortex

Contents

The read count ratio approach

Castel et al 2015 GenomeBiology

First part
Second part
Imprinting and allelic bias
Imprinted genes in the brain
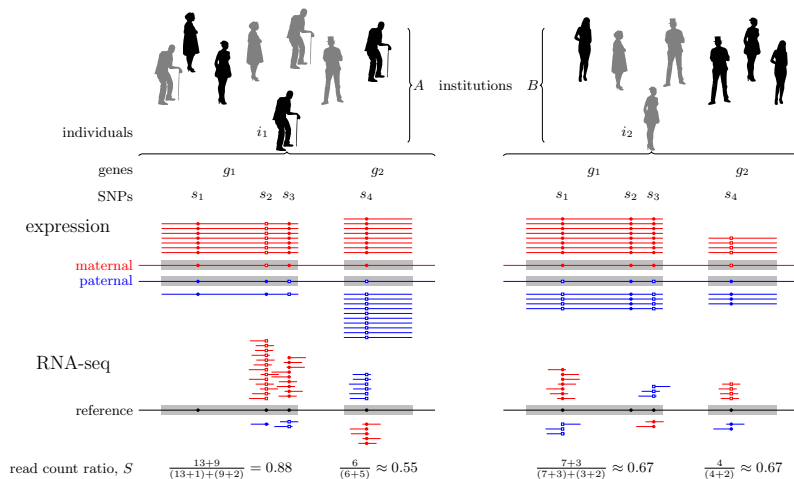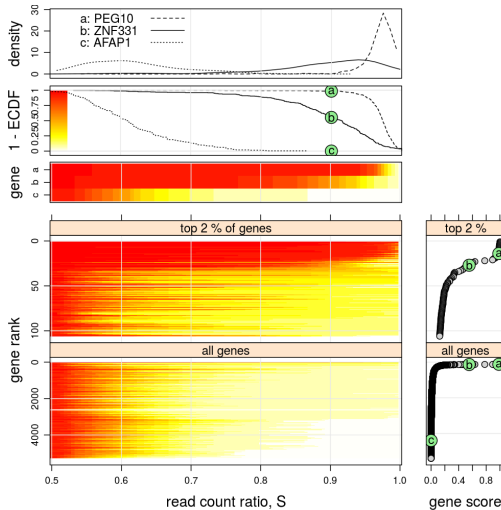
Our research study

data/project Common Mind Consortium

participants Ifat Keydar, Eva Xia, Menachem Fromer, Doug
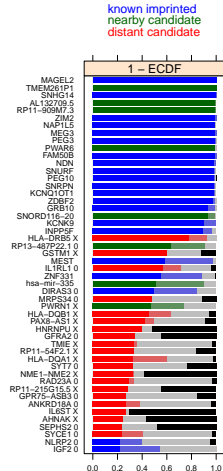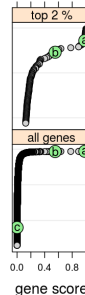Ruderfer, Ravi Sachinanandam, Andrew Chess

Study setup
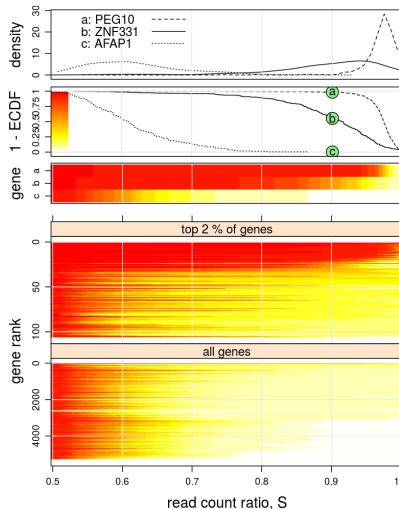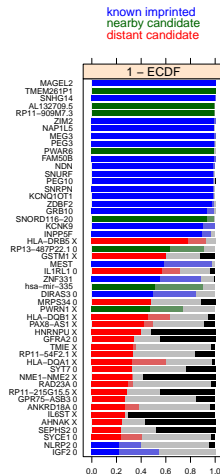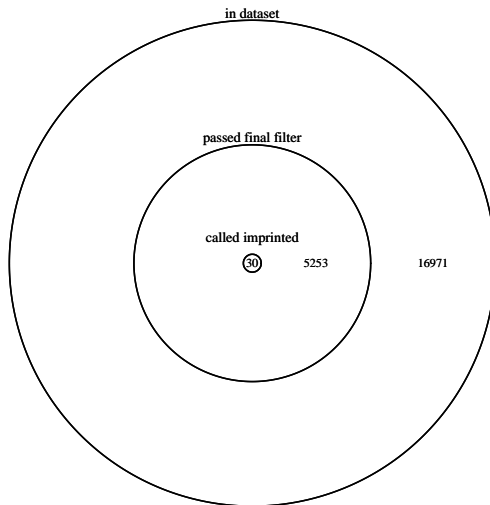
Distribution across individuals and genes

Called imprinted genes

Called imprinted genes

Questions and **Answers**

1. Imprinted genes in the DLPFC[2]
   - *How many?* **30 genes in $\approx \frac{1}{3}$ genome**
   - *Novelty?* **8 new imprinted genes**

2. Dependence of allelic bias...
   on age, psychiatric condition, genetics (ancestry), gender
   - *Signal vs noise?*
   - *Genes affected uniformly?*
   - *Most prominent effects?*

---

[2]dorsolateral prefrontal cortex
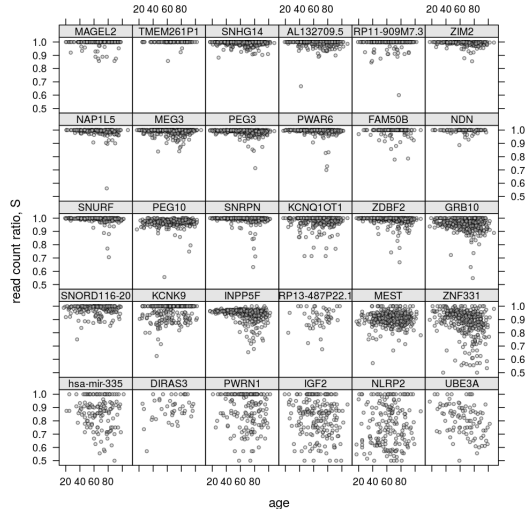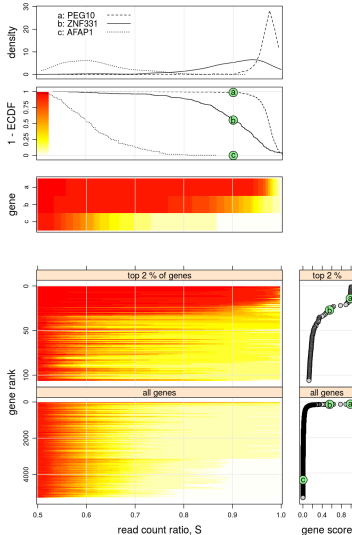
Contents

Explaining inter-individual variation

Multiple interdependent predictors

Inferring dependence using regression models

generalized linear models

$$
\begin{aligned}
\mathrm{E}[y_g] &= \mu_g = h^{-1}(X\beta_g) \\
y &= \mu_g + \varepsilon_{\mu_g}
\end{aligned}
$$
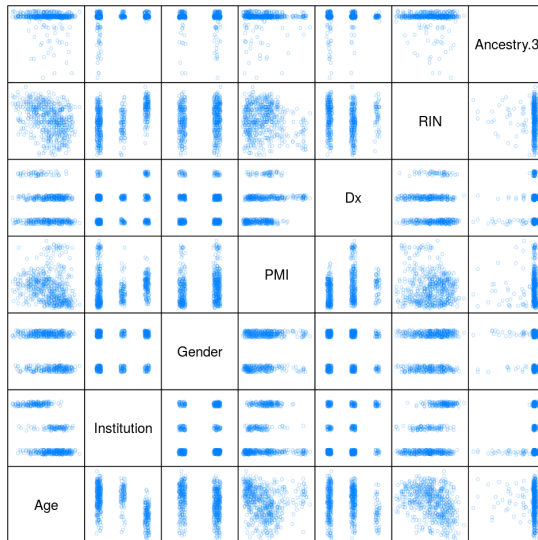
Inferring dependence using regression models

generalized linear models

$$\mathrm{E}[y_g] = \mu_g = h^{-1}(X\beta_g)$$
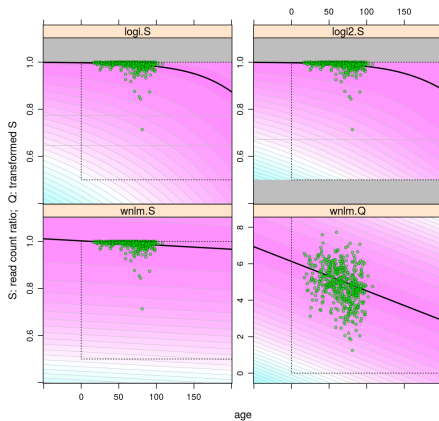$$y = \mu_g + \varepsilon_{\mu_g}$$

inference for gene g (PEG3)
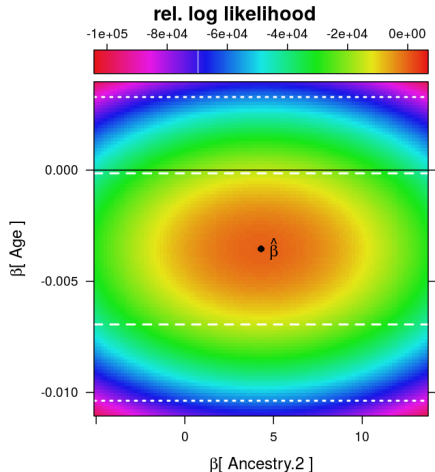
Selection among several models

predictive distributions
(simple regression, for demonstration only)
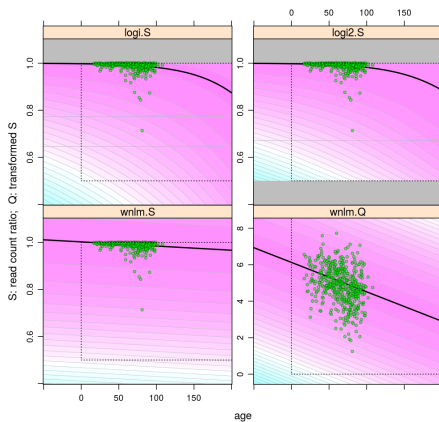
Selection among several models

predictive distributions
(simple regression, for demonstration only)
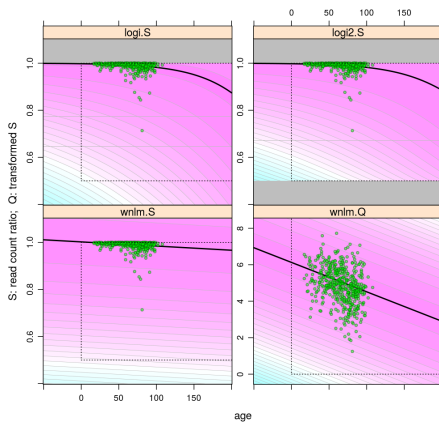

result:

AIC isn't useful in this case
bad fit may inflate likelihood for some models

Selection among several models

predictive distributions
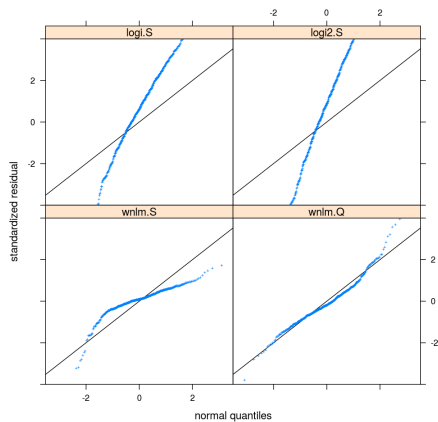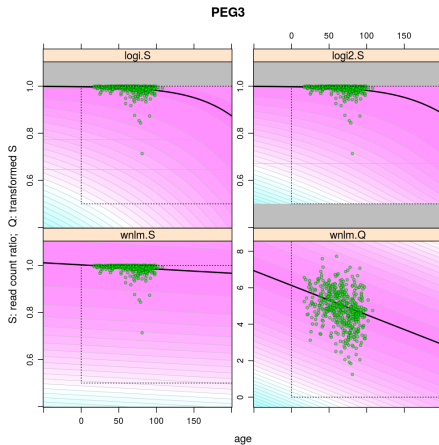(simple regression, for demonstration only)

normality of residuals
(multiple regression)

Selection among several models

predictive distributions
(simple regression, for demonstration only)

homogeneity of error
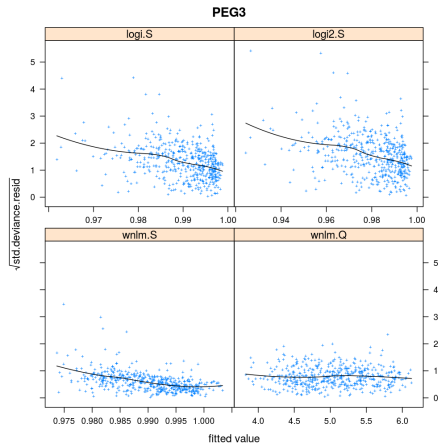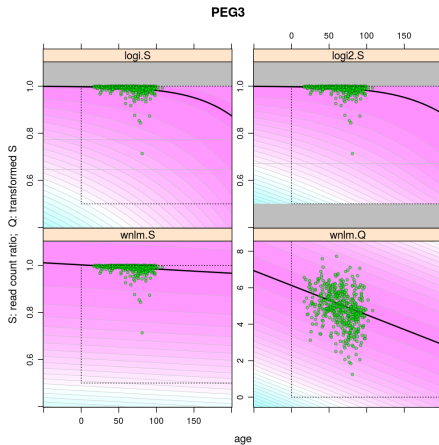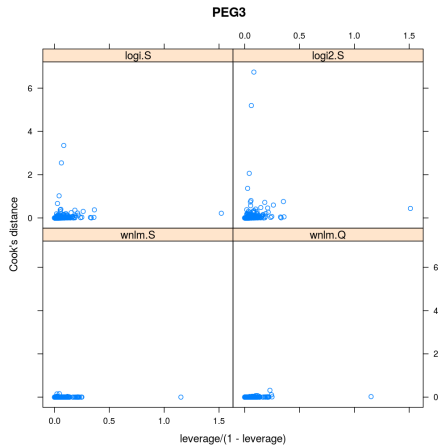(multiple regression)

Selection among several models

predictive distributions
(simple regression, for demonstration only)
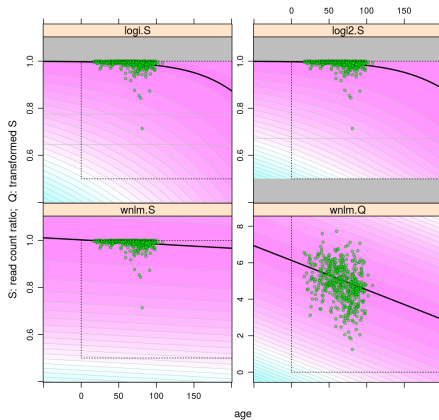
influence of outliers
(multiple regression)
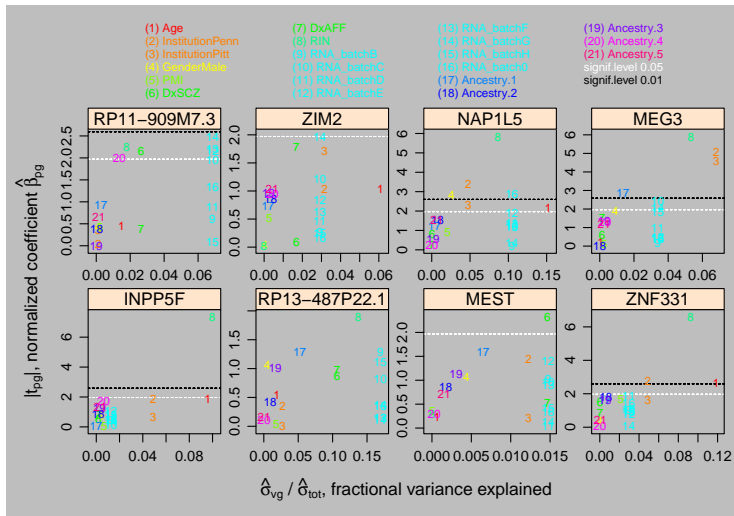
Selection among several models

predictive distributions
(simple regression, for demonstration only)

result:

wnlm.Q and unlm.Q fit the best

Explained variation of read count ratio

Genes g affected by one or more predictors p ($\beta_{pg} \neq 0$)

| Gene	Gene type	Chr	Coefficient	Known phenotype
ZDBF2	protein coding	2	Age, Ancestry.1	
NAP1L5	protein coding	4	GenderMale	
PEG10	protein coding	7	DxSCZ	
MEST	protein coding	7	DxSCZ	Silver-Russell syndrome
KCNK9	protein coding	8	Age	Birk-Barel mental retardation dysmorphism syndrome
INPP5F	protein coding	10	Age	cell motility; endocytic recycling
KCNQ1OT1	antisense	11	GenderMale	Beckwith-Wiedemann syn.; Isol. hemihyperplasia
MEG3	lincRNA	14	GenderMale	Mat/pat 14q32.2 hypermeth/microdel syndrome
RP11-909M7.3	lincRNA	14	DxSCZ	
AL132709.5	miRNA	14	Ancestry.1	
MAGEL2	protein coding	15	Age	Prader-Willi syn.; Schaaf-Yang syn.; Arthrogryposis
NDN	protein coding	15	GenderMale	Prader-Willi syndrome
PWRN1	lincRNA	15	Ancestry.1	Prader-Willi syndrome
UBE3A	protein coding	15	DxSCZ	Prader-Willi syn.; Angelman syn.; circadian rhythm
PEG3	protein coding	19	GenderMale	

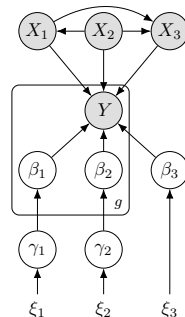# Different results under similarly well-fitting models
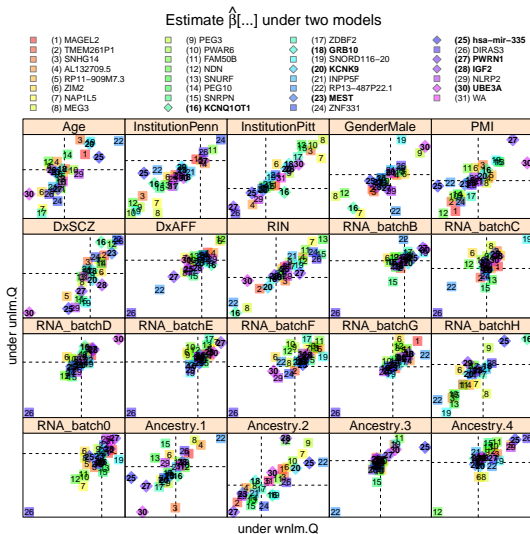
Effects appear interdependent
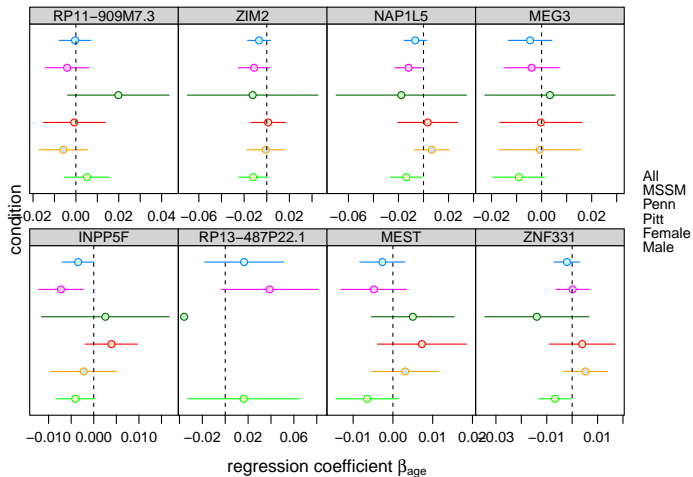
Multiple levels of variation
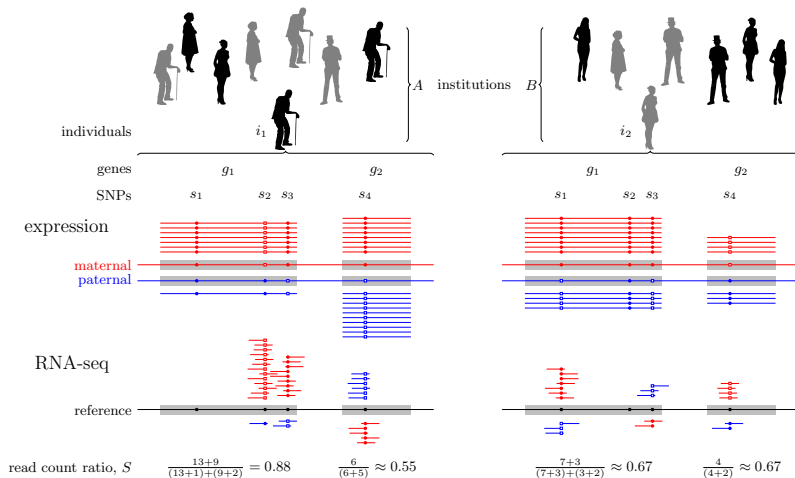
Better modeling framework?
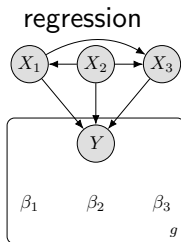
Questions and **Answers**

1. Imprinted genes in the DLPFC[2]
   - *How many?* **30 genes in $\approx \frac{1}{3}$ genome**
   - *Novelty?* **8 new imprinted genes**

2. Dependence of allelic bias...
   on age, psychiatric condition, genetics (ancestry), gender
   - *Signal vs noise?* **Poor; modeling challenge**
   - *Genes affected uniformly?* **No**
   - *Most prominent effects?* **genetics and age on a few genes**

---

[2]dorsolateral prefrontal cortex

Contents

Reanalyze dependence?

Is it worth? If yes…

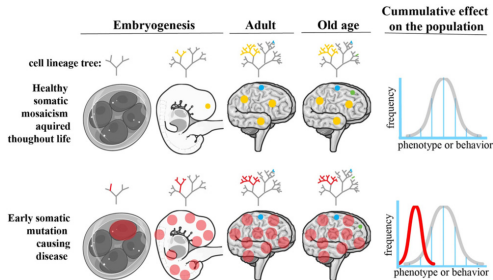1. add more data

2. more accurate read counts

3. find a better model

Reanalyze dependence?

Is it worth? If yes...

1. add more data
   - no solution for biased stats. approach
2. more accurate read counts
   - improve QC: RNA-seq + genotyping
3. find a better model
   - implement inference, validate

Brain Somatic Mosaicism
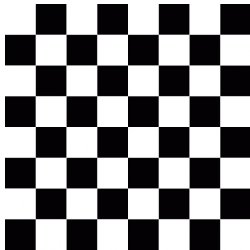

Paquola, Erwin, Gage 2016

challenges with somatic variants:

1. detection
   allelic fraction

2. prioritization
   multiple info

3. integration
   germline vars.

Thanks to

Chess lab

- Andy Chess (support)
- Chaggai Rosenbluh (feedback)
- Eva Xia
- Mehaa Bajaj

- Gabriel Hoffman
  (feedback, variancePartition)
- Ravi Sachinanandam
  (critical feedback)