

Identifying sample mix-ups in eQTL data

Karl Broman

Biostatistics & Medical Informatics, Univ. Wisconsin–Madison

`kbroman.org`

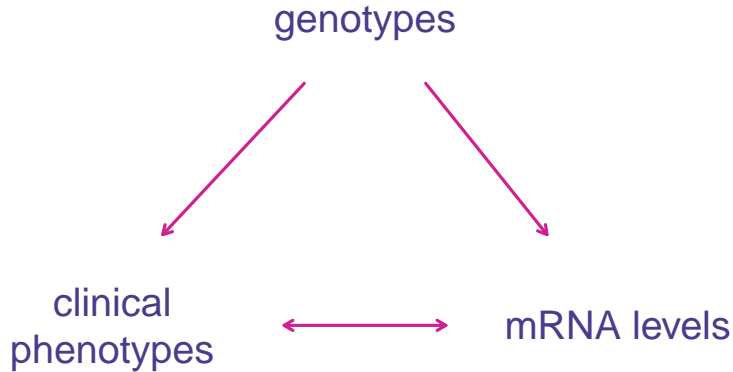
`github.com/kbroman`

`@kwbroman`

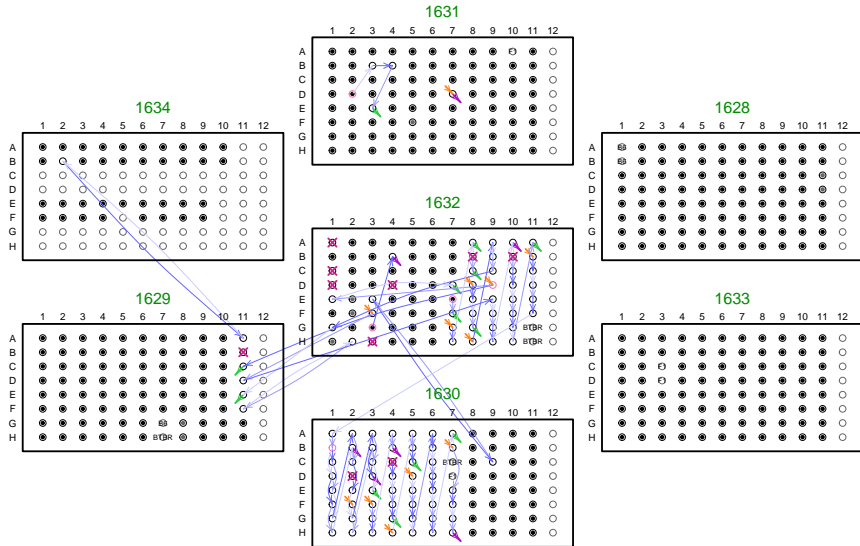
Slides: `kbroman.org/Talk_OSGA2021`



Associations in systems genetics

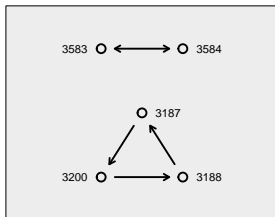


Sample mix-ups

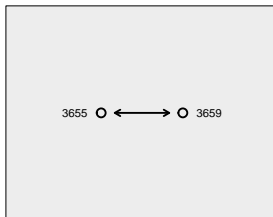


More sample mix-ups

adipose



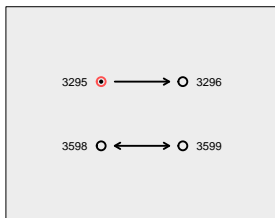
gastroc



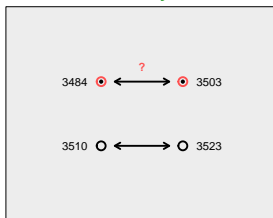
hypo



islet



kidney



liver

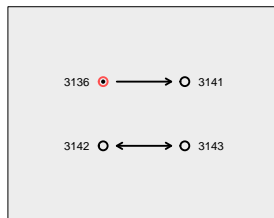


Table 2. *Cis*-eQTL mapping and sample mix-up identification results

Stud	Population	Sample-size	Initial <i>cis</i> -eQTLs	Mix-ups detected ^a <i>n</i> (%)	Sample-size after correction <i>n</i> (%)	<i>cis</i> -eQTLs after correction <i>n</i> (%)
Choy <i>et al.</i> (2008)	CHB+JP	87	138	20 (23)	79 (90)	418 (+203)
	CE	84	558		NA	NA
	YR	85	274	2 (2)	83 (97)	287 (+5)
Stranger <i>et al.</i> (2007)	CHB+JP	90	1511		NA	NA
	CE	90	903		NA	NA
	YR	90	663	1 (1)	89 (99)	667 (+1)
Zhang <i>et al.</i> (2009)	CE	87	2581		NA	NA
	YR	89	1454	2 (2)	89 (100)	1635 (+12)
Webster <i>et al.</i> (2009)	Brai	36	1284	16 (4)	356 (98)	1367 (+6)
Heinzen <i>et al.</i> (2008)	Brai	93	349		NA	NA
	PBMC	80	297		NA	NA

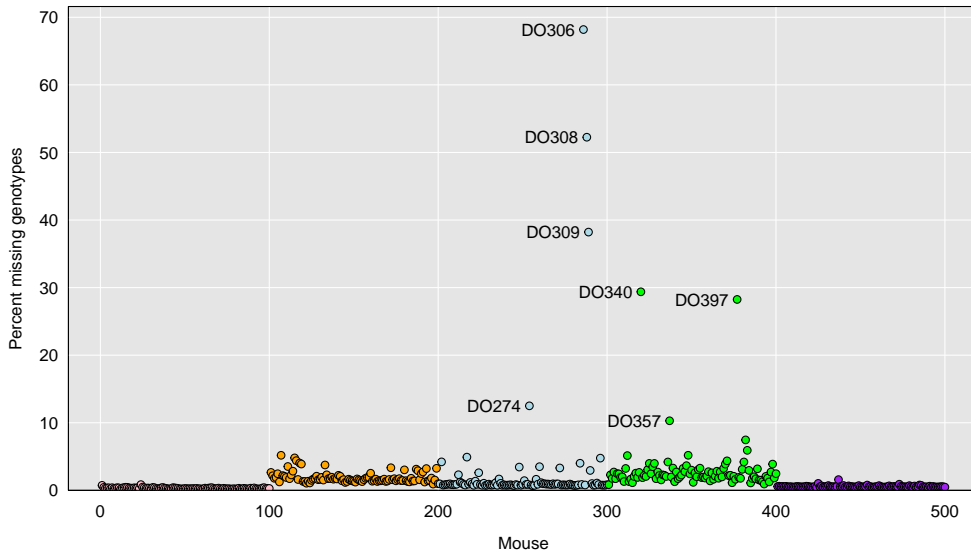
Outline

- ▶ Sample duplicates
- ▶ Sex verification
- ▶ mRNA \leftrightarrow protein
- ▶ mRNA \leftrightarrow DNA
- ▶ protein \leftrightarrow DNA

But first

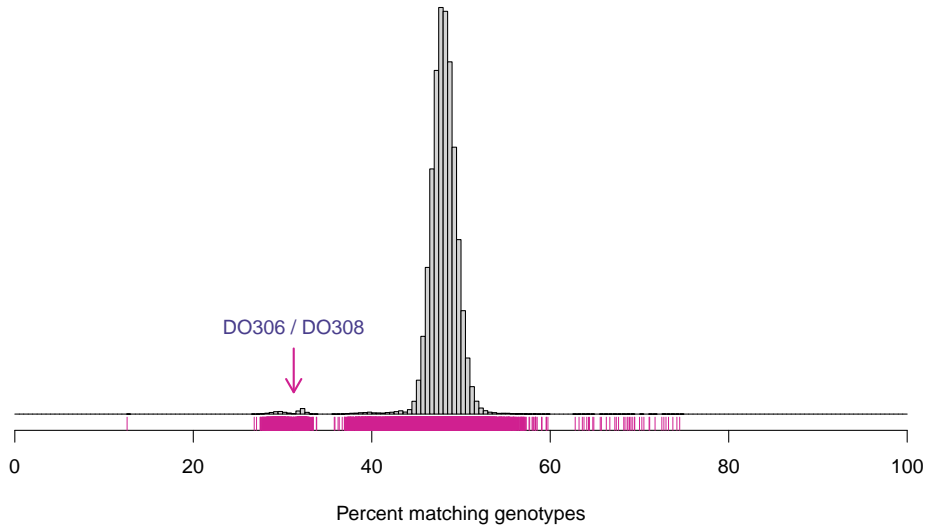
Missing Data

Percent missing genotypes



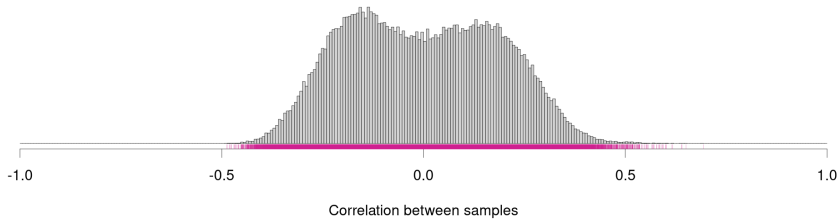
Sample duplicates

Percent matching genotypes

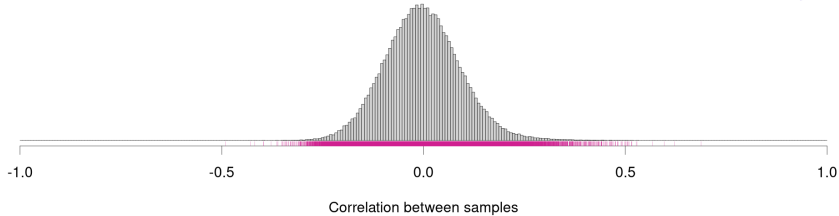


Correlation between mRNA samples

mRNAs

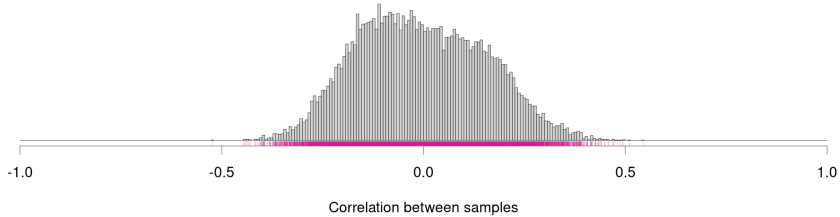


mRNAs
sex-adjusted

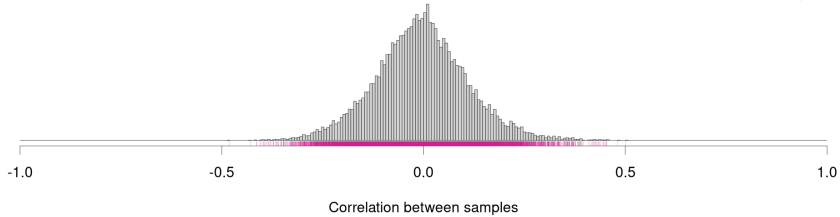


Correlation between protein samples

Proteins

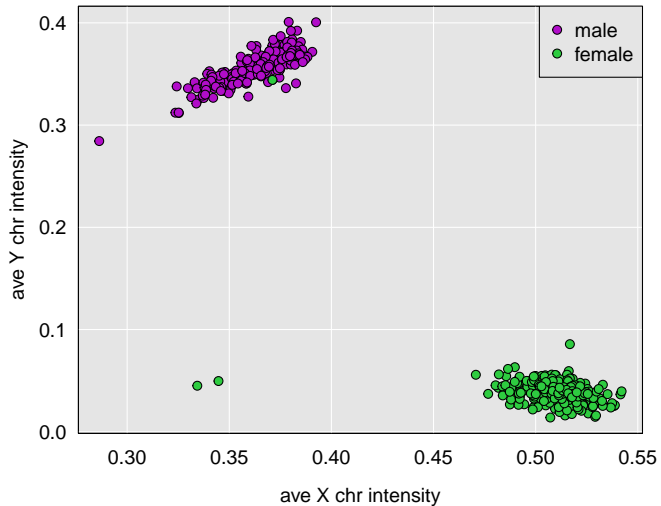


Proteins
sex-adjusted

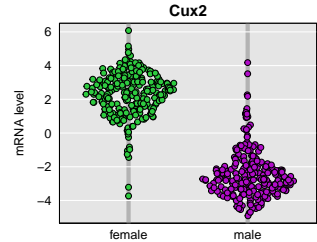
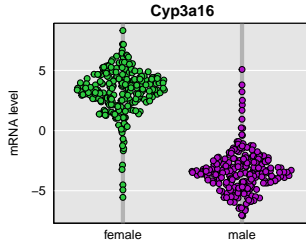
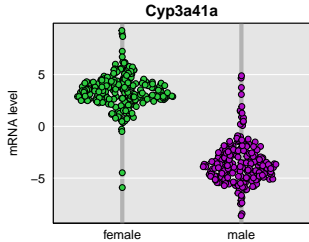
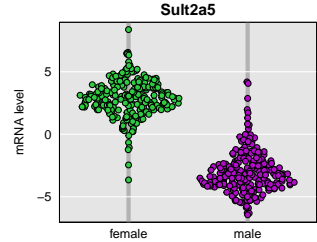
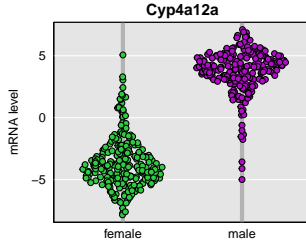
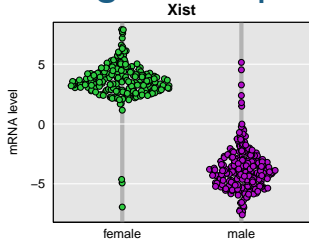


Sex verification

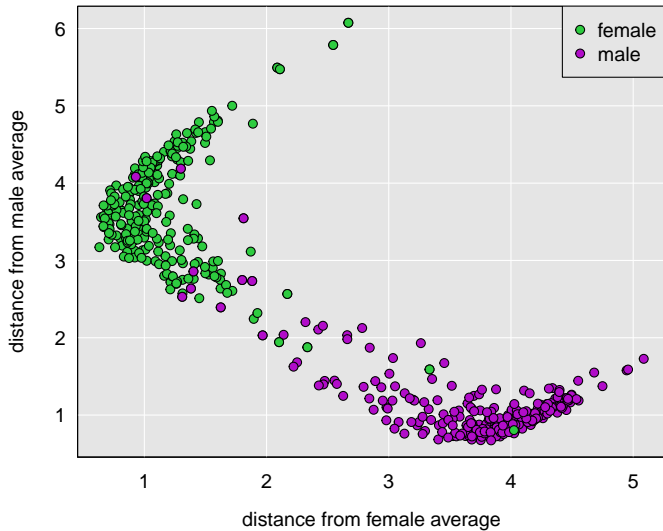
X and Y genotype dosage



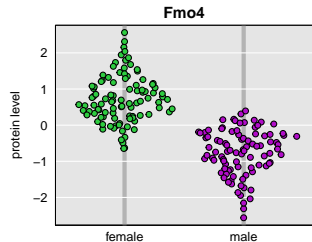
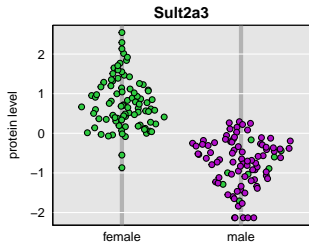
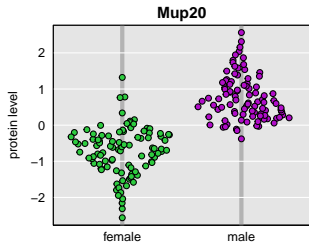
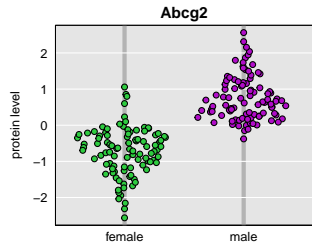
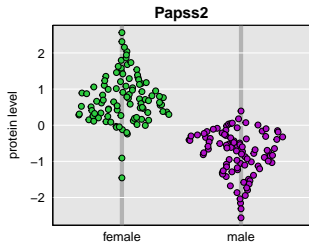
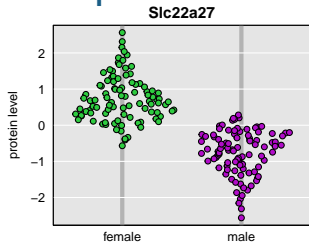
Sex and gene expression



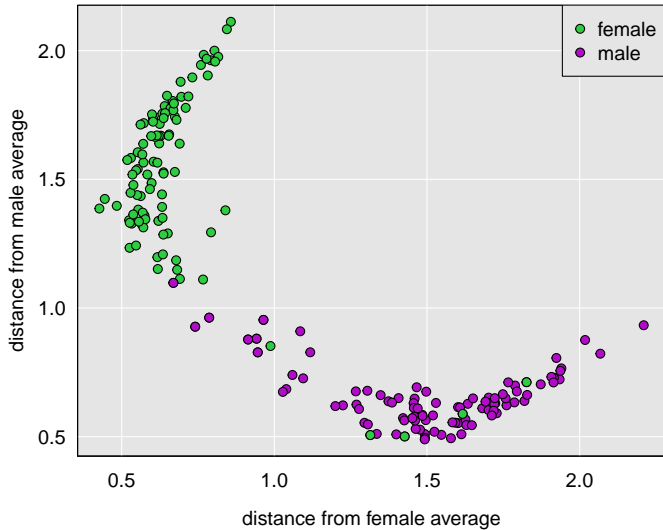
Sex and gene expression



Sex and proteins



Sex and proteins



mRNA \leftrightarrow protein

DNA \leftrightarrow mRNA

DNA \leftrightarrow protein

References

- ▶ Westra et al. (2011) MixupMapper: correcting sample mix-ups in genome-wide datasets increases power to detect small genetic effects. *Bioinformatics* 15:2104–2111 [doi:10.1093/bioinformatics/btr323](https://doi.org/10.1093/bioinformatics/btr323)
- ▶ Lynch et al (2012) Calling sample mix-ups in cancer population studies. *PLOS One* 7:e41815 [doi:10.1371/journal.pone.0041815](https://doi.org/10.1371/journal.pone.0041815)
- ▶ Broman et al. (2015) Identification and correction of sample mix-ups in expression genetic data: A case study. *G3 (Bethesda)* 5:2177–2186 [doi:10.1534/g3.115.019778](https://doi.org/10.1534/g3.115.019778)
- ▶ Broman et al. (2019) Cleaning genotype data from Diversity Outbred mice. *G3 (Bethesda)* 9:1571–1579 [doi:10.1534/g3.119.400165](https://doi.org/10.1534/g3.119.400165)

Slides: kbroman.org/Talk_OSGA2021



`kbroman.org`

`github.com/kbroman`

`@kwbroman`