

# X Chromosome Association Testing in HCHS/SOL

Caitlin McHugh, with Tim Thornton

Department of Biostatistics  
University of Washington

27 March 2015

# Outline

MLM-X: Mixed Model Association on the X Chromosome

Estimating Family Structure on the X Chromosome

Inferring Population Structure on the X Chromosome

Application to RBC Trait

## Mixed Model Association Mapping

- For association testing with an *autosomal* SNP, the following model has been proposed

$$Y = \beta_0 + \beta_1 \text{SNP} + g_A + \text{covariates} + \epsilon \quad (1)$$

where

$$g_A \sim MVN(0, \sigma_A^2 \Phi_A)$$

$$\epsilon \sim MVN(0, \sigma_\epsilon^2 \mathbb{I})$$

- For detecting association on the X chromosome, we propose the following **MLM-X** model

$$Y = \beta_0 + \beta_1 \text{SNP}_X + g_A + g_X + \text{covariates} + \epsilon \quad (2)$$

with an additional random component for polygenic effects on the X chromosome

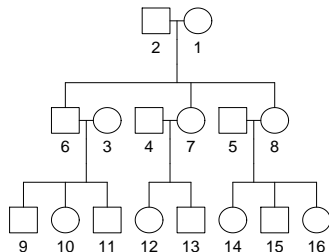
$$g_X \sim MVN(0, \sigma_X^2 \Phi_X)$$

# Sample Structure

- ▶ GWAS often have sample structure, including population stratification, family structure and/or cryptic relatedness.
- ▶ It is well known that failure to appropriately account for sample structure can lead to spurious association and reduced power.
- ▶ SNP data can be used to infer recent relatedness (family structure) and more distant relatedness (population structure).

## Genetic Relatedness on the X Chromosome: $\Phi_X$

- ▶ The X chromosome kinship coefficient between individuals  $i$  and  $j$ ,  $\Phi_{ij}^X$ , is defined as the probability of sampling one allele IBD at random from individual  $i$  and individual  $j$  on the X chromosome.
- ▶ *Note* for males there is no randomness in sampling, as there is only one allele at each location on the X chromosome.



		Autosomes	X Chromosome
	Self, Female	$\frac{1}{2}$	$\frac{1}{2}$
	Self, Male	$\frac{1}{2}$	1
	Mother-Daughter	$\frac{1}{4}$	$\frac{1}{4}$
	Mother-Son, Father-Daughter	$\frac{1}{4}$	$\frac{1}{2}$
	Father-Son	$\frac{1}{4}$	0
	Full sisters	$\frac{1}{4}$	$\frac{6}{16}$
	Full brothers	$\frac{1}{4}$	$\frac{1}{2}$
	Sister-Brother	$\frac{1}{4}$	$\frac{1}{4}$
Maternal	Aunt-Niece	$\frac{1}{8}$	$\frac{3}{16}$
	Aunt-Nephew	$\frac{1}{8}$	$\frac{6}{16}$
	Uncle-Niece	$\frac{1}{8}$	$\frac{1}{8}$
	Uncle-Nephew	$\frac{1}{8}$	$\frac{1}{4}$
	Grandma-Granddaughter	$\frac{1}{8}$	$\frac{1}{8}$
	Grandma-Grandson	$\frac{1}{8}$	$\frac{1}{4}$
	Grandpa-Granddaughter	$\frac{1}{8}$	$\frac{1}{4}$
	Grandpa-Grandson	$\frac{1}{8}$	$\frac{1}{2}$
Paternal	Aunt-Niece	$\frac{1}{8}$	$\frac{1}{8}$
	Aunt-Nephew	$\frac{1}{8}$	0
	Uncle-Niece	$\frac{1}{8}$	0
	Uncle-Nephew	$\frac{1}{8}$	0
	Grandma-Granddaughter	$\frac{1}{8}$	$\frac{1}{4}$
	Grandma-Grandson	$\frac{1}{8}$	0
	Grandpa-Granddaughter	$\frac{1}{8}$	0
	Grandpa-Grandson	$\frac{1}{8}$	0

## Estimating $\Phi_X$

- We can estimate  $\Phi_X$  using the following GRM equations:

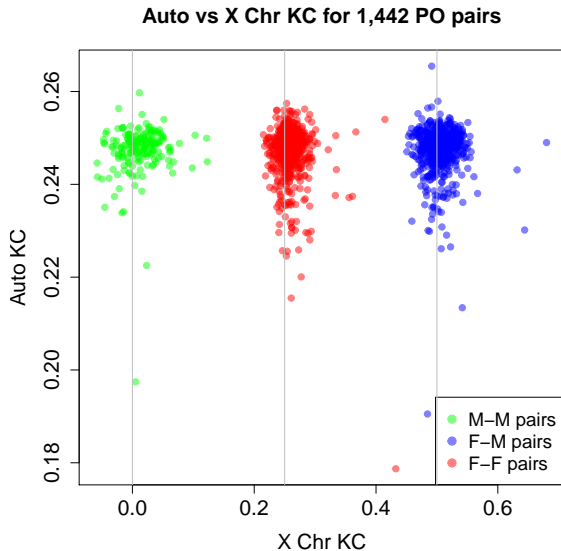
$$GR_{FF} = \frac{1}{N} \frac{\sum_{i=1}^N (X_{iF} - 2p_i)(X_{iF} - 2p_i)}{\sum_{i=1}^N 2p_i(1 - p_i)}$$

$$GR_{MM} = \frac{1}{N} \frac{\sum_{i=1}^N (X_{iM} - p_i)(X_{iM} - p_i)}{\sum_{i=1}^N p_i(1 - p_i)}$$

$$GR_{MF} = \frac{1}{N} \frac{\sum_{i=1}^N (X_{iM} - p_i)(X_{iF} - 2p_i)}{\sum_{i=1}^N \sqrt{2}p_i(1 - p_i)}$$

where  $F$  indicates a female and  $M$  is a male.

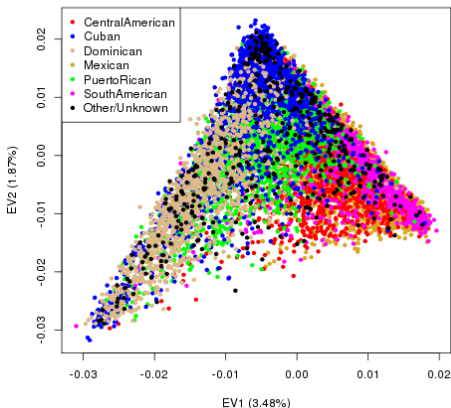
# HCHS/SOL Samples, For Example





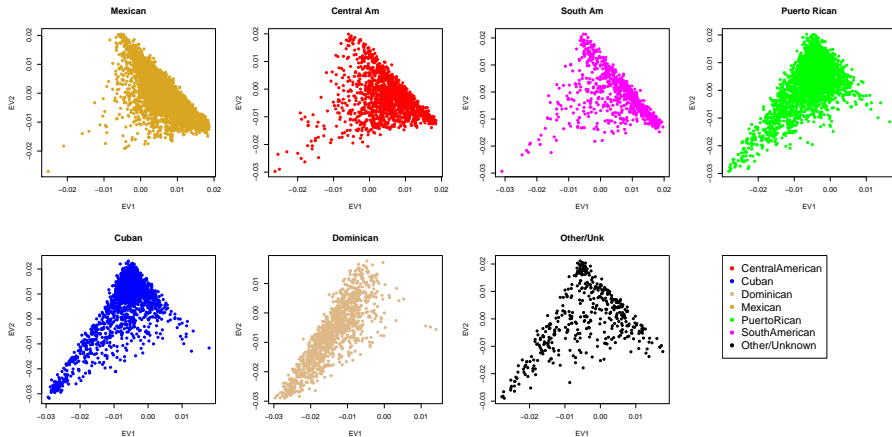
# Population Structure on the X Chromosome

We performed PCA using 3,600 LD pruned X chromosome SNPs.



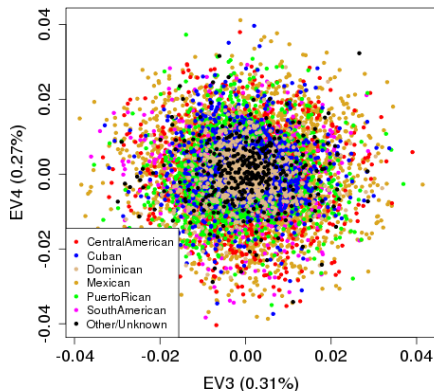
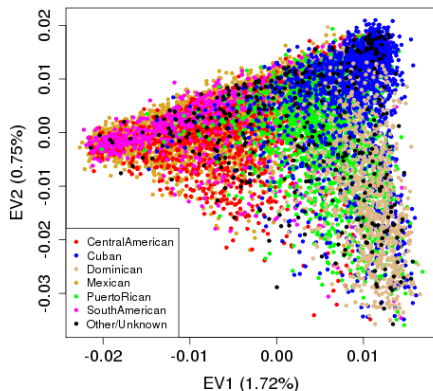
Can we control for this structure by simply adjusting for the autosomal PCs?

# Population Structure on the X Chromosome



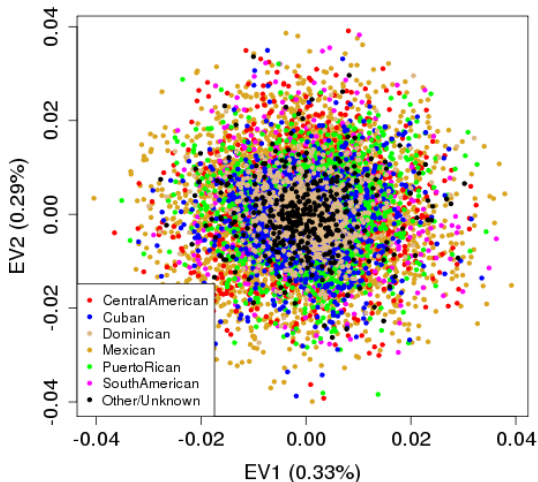
# Population Structure on the X Chromosome, Beyond the Autosomes

We estimated PCs on the X chromosome after adjusting for autosomal PCs 1-5.

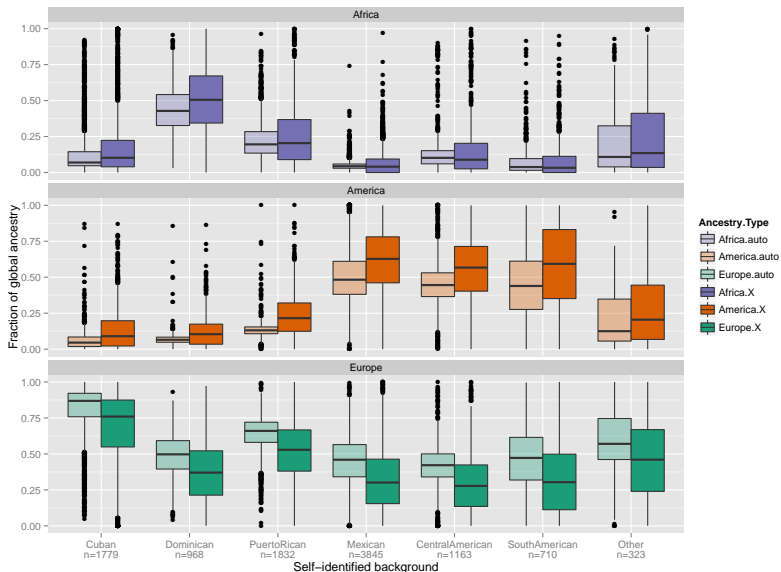


## Population Structure on the X Chromosome?

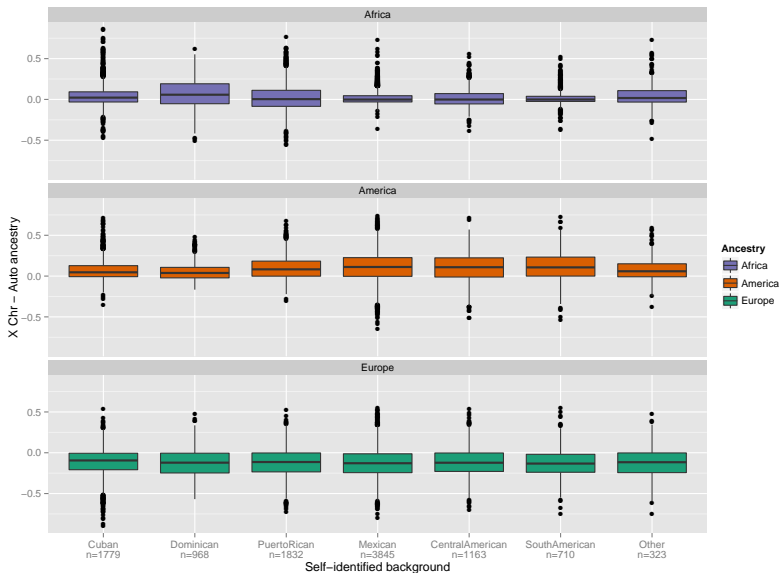
We estimated PCs on the X chromosome after adjusting for X chromosome PCs 1-2.



# Global Ancestry Across the Genome



# Global Ancestry Difference Between X Chr & Autosomes



# Red Blood Cell Count

- ▶ We performed a GWAS with the red blood cell count (RBC) trait.
- ▶ RBC was previously found to be associated with variants in gene G6PD on Xq28 in African Americans.
- ▶ We found the Xq28 region to be associated with RBC in the HCHS/SOL samples, when applying the autosomal mixed model and the MLM-X model.

## 'Autosomal Model'<sup>1</sup>

We include all individuals with non-missing outcome and covariates, exclude Asian outliers identified with PCA and additionally exclude samples with

- ▶ blood/lymph malignant tumor
- ▶ bone cancer
- ▶ pregnancy
- ▶ chronic kidney disease
- ▶ chemotherapy
- ▶ % blasts >5
- ▶ % immature granulocytes >5

Fixed effect covariates included are sex, age, center, and autosomal ancestry eigenvectors 1-5.

We include random effects of block group, household and autosomal kinship.

---

<sup>1</sup>these are as described in the working group previously

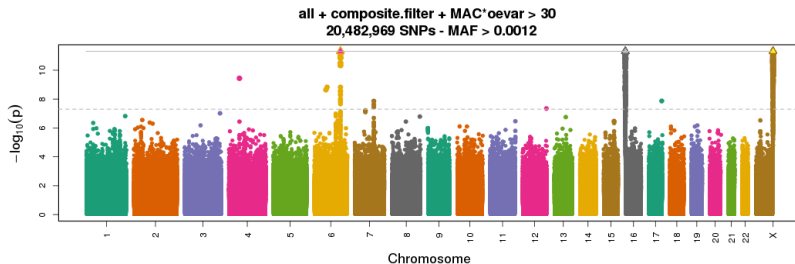
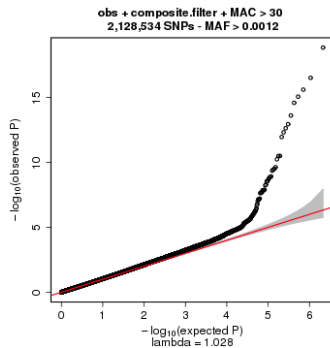


## Previous Variance Component Estimates, n=12,502

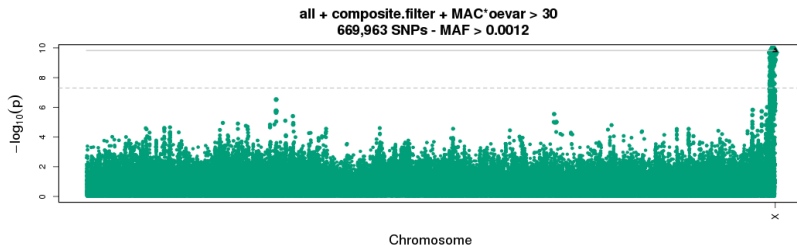
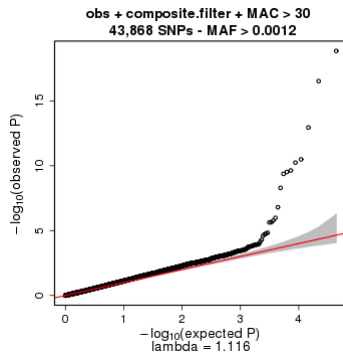
	estimate	95% CI
block group	0.00363	(-0.00177, 0.00903)
household	0.04945	(0.02113, 0.07777)
autosomal kinship	0.28473	(0.23215, 0.33732)
environment	0.66218	(0.60989, 0.71448)

**Table:** Estimate (95% CI) of the proportion variance for each of the components.

# Previous Genome-Wide Assoc Test Results, n=12,502



# Previous X Chr Assoc Test Results, n=12,502



## MLM-X Model

We include all individuals with non-missing outcome and covariates, exclude Asian outliers identified with PCA and additionally exclude samples with

- ▶ blood/lymph malignant tumor
- ▶ bone cancer
- ▶ pregnancy
- ▶ chronic kidney disease
- ▶ chemotherapy
- ▶ % blasts >5
- ▶ % immature granulocytes >5
- ▶ **an X chromosome anomaly**

Fixed effect covariates included are sex, age, center, autosomal ancestry eigenvectors 1-5, and **X chromosome ancestry eigenvectors 1-2**.

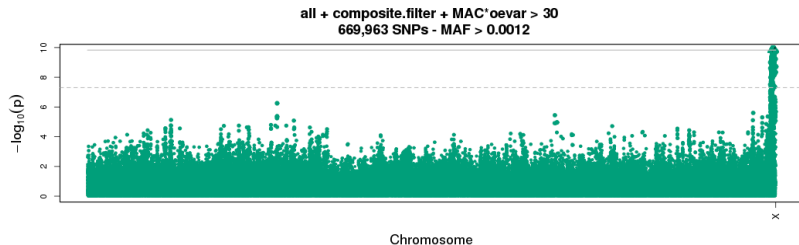
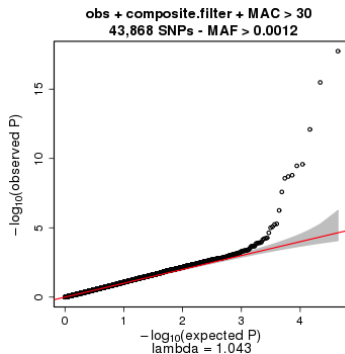
We include random effects of block group, household, autosomal kinship and **X chromosome kinship**.

## MLM-X Variance Component Estimates, n=12,488

	MLM-X	Autosomal
block group	0.00396 (-0.0016, 0.0095)	0.00363 (-0.0018, 0.0090)
household	0.04950 (0.0209, 0.0781)	0.04945 (0.0211, 0.0778)
autosomal kinship	0.28453 (0.2316, 0.3375)	0.28473 (0.2322, 0.3373)
X kinship	0.02935 (0.0137, 0.0450)	
environment	0.63266 (0.5783, 0.6870)	0.66218 (0.6099, 0.7145)

**Table:** Estimate (95% CI) of the proportion variance for each of the components.

# MLM-X Assoc Test Results, n=12,488



## Index SNP rs1050828

	Effect Size (SE)	Stat	p-value
autosomal	0.1314 (0.0145)	81.9377	1.40447e-19
MLM-X	0.1300 (0.0148)	76.8729	1.82321e-18