lme4qtl: an efficient and flexible tool for QTL mapping in related individuals

HOSPITAL DE LA SANTA CREU I SANT PAU
UNIVERSITAT AUTÒNOMA DE BARCELONA

Andrey Ziyatdinov[1], Helena Brunel[1], Miquel Vázquez-Santiago[2], Angel Martinez-Perez[1], José Manuel Soria[1]

(1) Unitat de Genomica de Malalties Complexes. Institut d'Investigació Biomédica Sant Pau (IIB-Sant Pau). Barcelona. Spain.
(2) Unitat de Trombosis i Hemostasia. Hospital de la Santa Creu i Sant Pau. Barcelona. Spain

About lme4qtl

lme4 (Bates et al.) provided linear mixed models (LMMs) in R since 2002, while lme4qtl extends it with random effects defined via the covariance matrices, e.g. the double kinship matrix.

Use cases of lme4qtl
- Regress out experimental noise (e.g. gene expression data)
- Introduce custom similarity matrices (e.g. pathway s) into LMMs
- Bootstrap parameters (lme4); exact significance tests (RLRsim)

Failures when using tools other than lme4qtl
- Get residuals of a model using SOLAR
- Run logistic regression for a binary trait using GEMMA
- Include twins in the study using pedigreemm

Features

Features of lme4 inherited by lme4qtl
- Allows arbitrarily many nested and crossed random effects
- Efficient for large data sets by using the Eigen package
- Fits generalized LMMs and nonlinear LMMs
  - Laplace approx. or adaptive Gauss-Hermite quadrature
  - GLMMs allow user-defined families and link functions

Features implemented in lme4qtl
- Cholesky decomposition of the covariance matrix and further update of the incidence matrix (Harville and Callanan)
- The nearest positive definite of a real symmetric matrix estimated for semidefinite covariance matrices (Higham)
- Constraints on the model parameters for the inference

Binary trait

Thrombosis is a common complex disease. The major determinants of thrombosis include both environmental and genetic factors.

- Enironmental: age & gender
- Genetic: ABO blood group system
  - Group O is protector
- The heritability is 60% (Souto et al.)
- The prevalence is between 0.2% & 2%

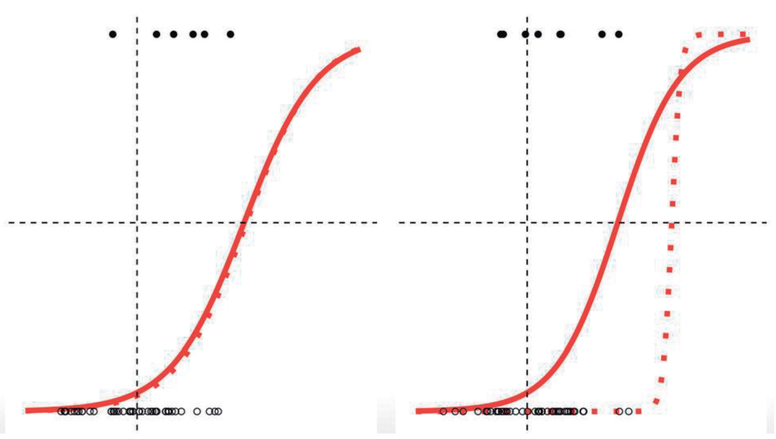
```
K <- 0.02
dat <- mutate(phen2, offset = -qnorm(1 - K))

m <- relmatGlmer(Throm ~ -1 + AGEsc + SEXfnum + ABOf3num + (1|ID), dat,
  offset = offset, relmat = list(ID = dkin2), family = binomial(probit))
```

```
summaryCoef(m)  # offset = -qnorm(1 - K) = -2.053749
```

```
##           Estimate Std. Error z value Pr(>|z|)
## AGEsc       0.8898     0.1048   8.491   <2e-16 ***
## SEXfnum     0.2728     0.1671   1.632   0.1026
## ABOf3num   -0.3037     0.1603  -1.894   0.0582 .
```

Modeling prevalence?

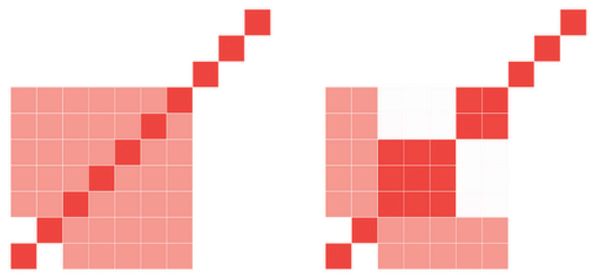
- Left: GAIT2 (118 cases *vs*. 817 controls)
- Right: GAIT1 (53 cases *vs*. 340 controls)
- Solid red line: prevalence is fixed
- Dashed red line: prevalence is free



Links

- Repository github.com/variani/lme4qtl
- Presentation lives here bit.ly/1UiTZvQ

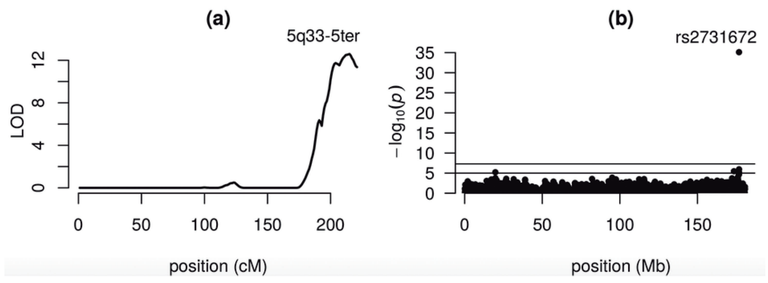
Combined Linkage & Assoc.



(*Soria et al.*) showed that a locus in the F12 gene influences both

- Coagulation Factor XII (FXII) activity
- Susceptibility to thrombosis

Figure below shows (a) linkage and (b) association mappings on Chromosome 5 for Factor FXII in the GAIT1 sample.



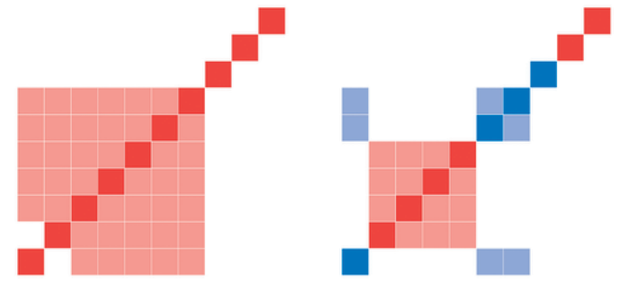
c46t is the key player

```
dat <- mutate(subset(phen, !is.na(c46t)), IBDID = ID)

m <- relmatLmer(FXII ~ c46t + (1|ID) + (1|IBDID), dat,
  relmat = list(ID = dkin, IBDID = mibd))

m0 <- update(m, . ~ . - c46t - (1|IBDID))

m1 <- update(m, . ~ . - c46t)
m2 <- update(m, . ~ . - (1|IBDID))
```

```
anova(m, m0, m1, m2)
```

```
## refitting model(s) with ML (instead of REML)

## Data: dat
## Models:
## m0: FXII ~ (1 | ID)
## m1: FXII ~ (1 | ID) + (1 | IBDID)
## m2: FXII ~ c46t + (1 | ID)
## m: FXII ~ c46t + (1 | ID) + (1 | IBDID)
##    Df    AIC    BIC  logLik deviance    Chisq Chi Df Pr(>Chisq)
## m0  3 3739.8 3751.6 -1866.9   3733.8
## m1  4 3698.8 3714.6 -1845.4   3690.8  42.9508      1  5.614e-11 ***
## m2  4 3592.7 3608.5 -1792.4   3584.7 106.0915      0  < 2.2e-16 ***
## m   5 3592.0 3611.8 -1791.0   3582.0   2.7186      1    0.09918 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Gene x Env. Interaction



Body mass index (bmi) is known to depend on age and gender. Do genetic components also depend on gender?

- Sex-specificity model shows
- P-value for the env. component: 0.03
- P-value for the gen. Component: 0.09

(anova test with df = 2 was used)

```
# Common polygenic model
m0 <- relmatLmer(
  BMI ~ AGEsc + AGEsc2 + SEXf + (1|HHID) + (1|ID),
  phen, relmat = list(ID = dkin))

# Sex-specificity only in the residual variance
m1 <- relmatLmer(
  BMI ~ AGEsc + AGEsc2 + SEXf + (1|ID) + (0 + SEXf|RID) + (1|HHID),
  phen, relmat = list(ID = dkin),
  weights = rep(1e10, nrow(phen)), vcControl = list(rho0 = list(rid = 3)))

# Sex-specificity in both polygenic and residual variances
m2 <- relmatLmer(
  BMI ~ AGEsc + AGEsc2 + SEXf + (0 + SEXf|ID) + (0 + SEXf|RID) + (1|HHID),
  phen, relmat = list(ID = dkin),
  weights = rep(1e10, nrow(phen)), vcControl = list(rho0 = list(rid = 5)))
```

lme4qtl is efficient

GWAS on GAIT2
- N = 934 individuals, M = 10M markers
- Server with RAM 128G, 64 CPU x 2.3G
- SOLAR 7.6.6 (stable on March, 2015)

Model	SOLAR (days)	lme4qtl (days)
`APTT ~ 1 + (1|ID)`	1.2	1.6
`APTT ~ AGE + SEX + (1|ID)`	1.6	1.6
`APTT ~ 1 + (1|HHID) + (1|ID)`	5.6	1.7
`APTT ~ AGE+ SEX + (1|HHID) + (1|ID)`	8.2	1.7

Andrey Ziyatdinov: aziyatdinov@santpau.cat.