MLE for Individual Ancestries Population Covariances and Selection

PhD Student: Jade Y. Cheng
Institution: Bioinformatics Research Centre
Department of Computer Science · Aarhus University

Supervisor: Rasmus Nielsen
Institution: Center for Theoretical Evolutionary Genomics
University of California Berkeley

Selection Study

$$\ln\left[P_1\left(Q,\ F\right)\right] = \sum_i^I \sum_j^J \left\{ g_{ij} \cdot \ln\left[\sum_k^K q_{ik} \cdot f_{kj}\right] + (2 - g_{ij}) \cdot \ln\left[\sum_k^K q_{ik} \cdot (1 - f_{kj})\right] \right\}$$

$$\ln\left[P_2\left(F\right)\right]$$

$$= \ln\left\{ \prod_j^J \left[\frac{1}{\sqrt{|2\pi c_j \Omega'|}} \exp\left(-\frac{1}{2} \cdot f_j'^T \cdot (c_j \Omega')^{-1} \cdot f_j' \right) \right] \right\}$$

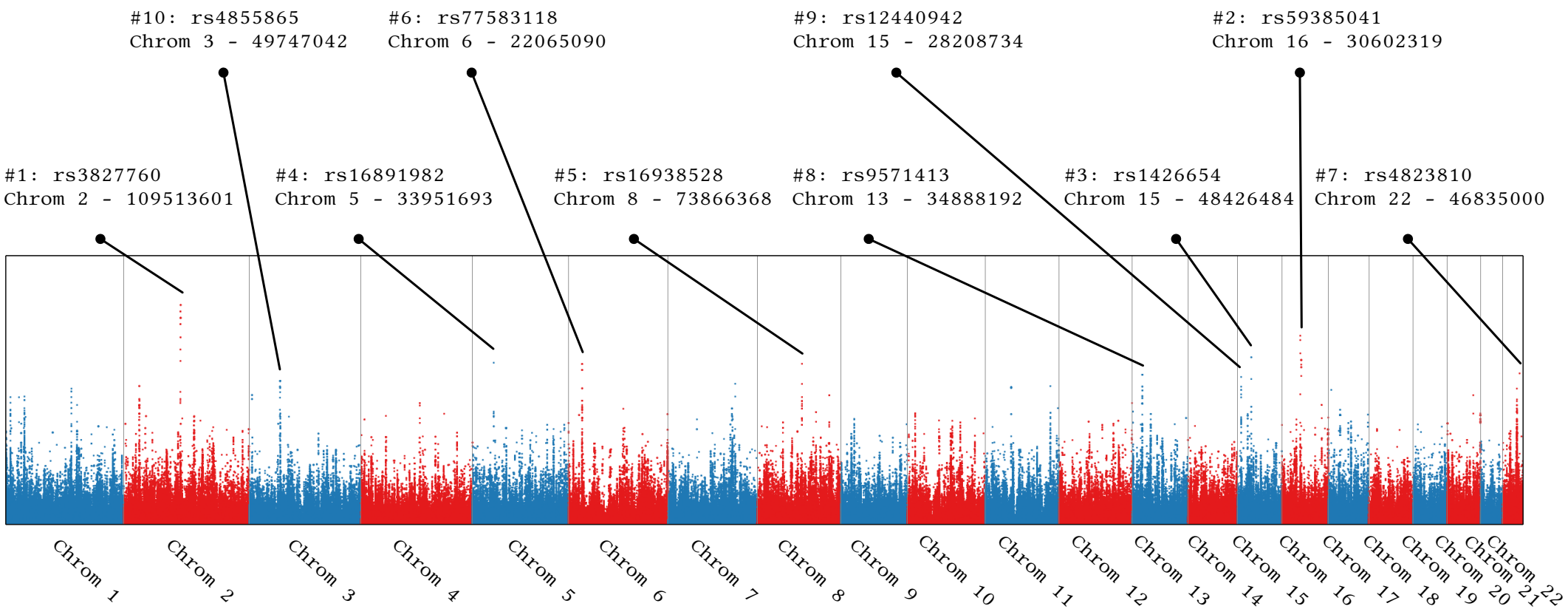
where $c_j = \mu_j\left(1 - \mu_j\right)$

$$f_j' = f_j - f_{j_0}.$$

SELECTION

Obtain the full genotype dataset G with N markers and M samples
Sample N' markers with respect to LD (N' > 100,000) to form G'
QPAS over ln(P_1) using G'
  Produce admixture proportions Q' of size M by K
  Produce allele frequencies F' of size K by N'
Nelder-Mead over ln(P_2) using F'
  Produce variance covariance matrix Ω'
QPAS over ln(P_1) using G while fixing Q'
  Produce allele frequencies F of size K by N
Repeat for each marker in F
  Set l_{ratio} to zero
  Repeat for each α in a range of an even interval starting from 1.0
    Set l_{new} to ln(P_2) calculated for this marker using α × Ω'
    Set l_{old} to ln(P_2) calculated for this marker using Ω'
    If 2 × (l_{new} - l_{old}) is greater than l_{ratio}
      Set l_{ratio} to 2 × (l_{new} - l_{old})
  End Repeat
  Emit l_{ratio}
End Repeat

Selection Study



Covariance Scan: Australian, English, Han, and Yoruba

Selection Study

hair thickness and curliness

earwax moisture underarm odor

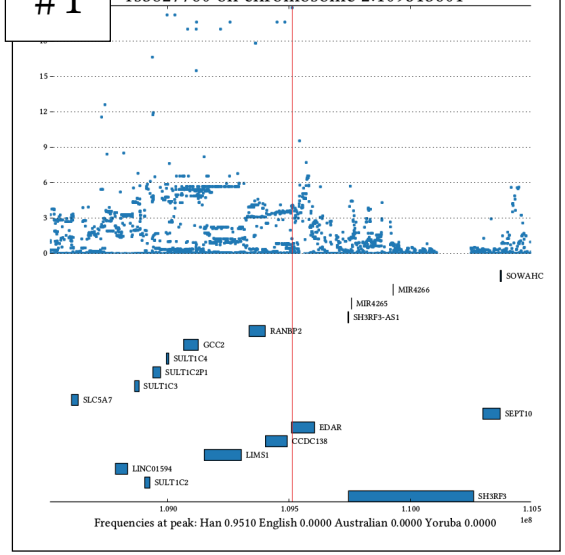
skin pigmentation

skin pigmentation

'maxDrinks' analcohol consumption related

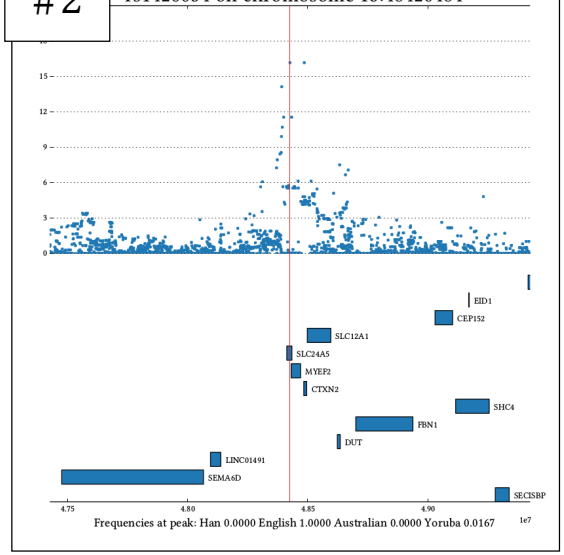
a melanoma tumor repressor

#1 rs3827760 on chromosome 2:109513601



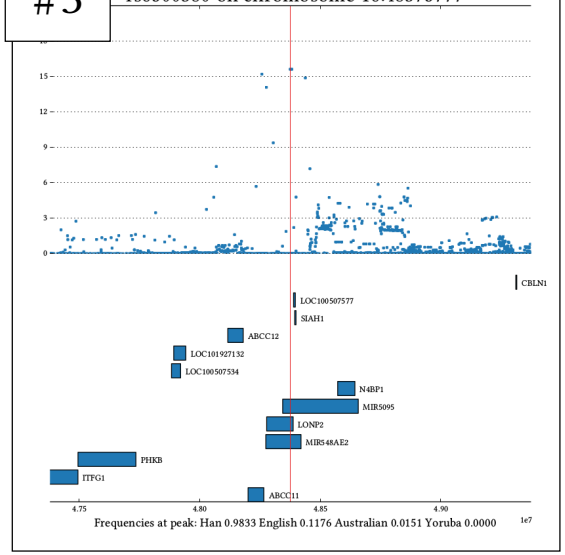
Frequencies at peak: Han 0.9510 English 0.0000 Australian 0.0000 Yoruba 0.0000

#2 rs1426654 on chromosome 15:48426484



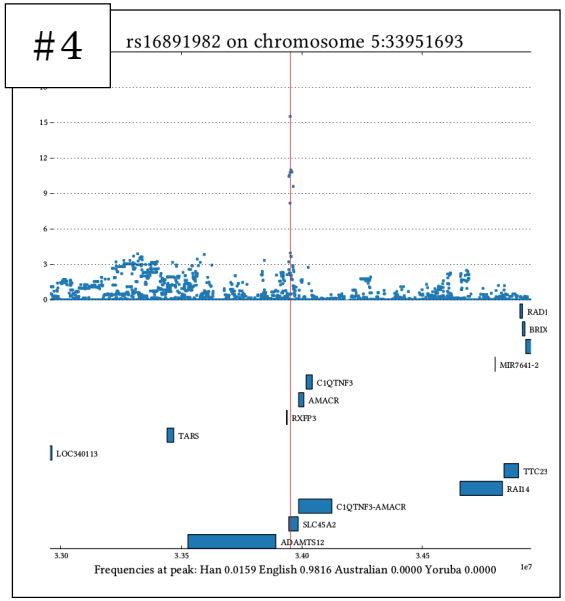
Frequencies at peak: Han 0.0000 English 1.0000 Australian 0.0000 Yoruba 0.0167

#3 rs6500380 on chromosome 16:48375777



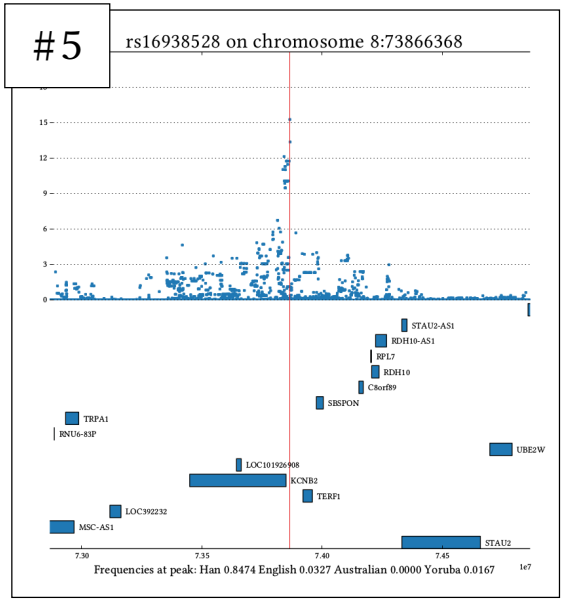
Frequencies at peak: Han 0.9833 English 0.1176 Australian 0.0151 Yoruba 0.0000

#4 rs16891982 on chromosome 5:33951693



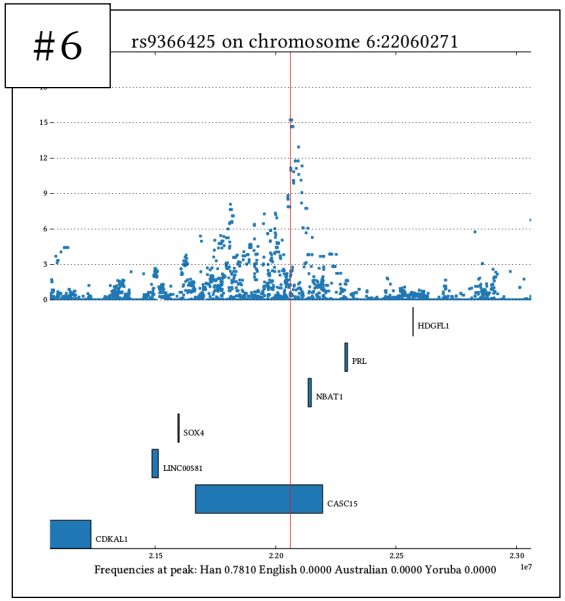
Frequencies at peak: Han 0.0159 English 0.9816 Australian 0.0000 Yoruba 0.0000

#5 rs16938528 on chromosome 8:73866368



Frequencies at peak: Han 0.8474 English 0.0327 Australian 0.0000 Yoruba 0.0167

#6 rs9366425 on chromosome 6:22060271



Frequencies at peak: Han 0.7810 English 0.0000 Australian 0.0000 Yoruba 0.0000

Selection Study

neural tube defect and hair follicles

skin pigmentation

Intergenic

Intergenic

insulin dependent regulation of glucose

taste cells in the mouth