

Sujet contrôle continu

Université Grenoble Alpes - Master Statistique - UE Biostatistique

Florent Chuffart & Magali Richard

2018-11-19

Sujet

On s'intéresse à la méthylation de l'ADN sur différents tissus humains (187 échantillons). Pour certains échantillons, le genre est manquant. Utilisez les valeurs de méthylation pour **imputer le sexe** des échantillons lorsqu'il n'est pas connu.

Veuillez restituer votre travail sous le format d'un PDF de deux pages (graphiques compris) en décrivant la méthode que vous avez choisie, en expliquant les choix que vous avez effectués et en analysant vos résultats.

Ce travail est à renvoyer par e-mail avant le mardi 27 novembre 2018 (minuit) à :

- magali.richard@univ-grenoble-alpes.fr & florent.chuffart@univ-grenoble-alpes.fr

Données

Chargement des données

```
install.packages("devtools")
devtools::install_github("fchuffart/methdbr")

d = methdbr::methdbr_d # A matrix of 187 x 137595 beta values (methylation of CpG probes)
e = methdbr::methdbr_e # A data frame that describes the 187 samples.
pf = methdbr::methdbr_pf # A data frame that describes the 137595 CpG probes.

table(e$sex, useNA = "ifany") # NA corresponds to missing sex data

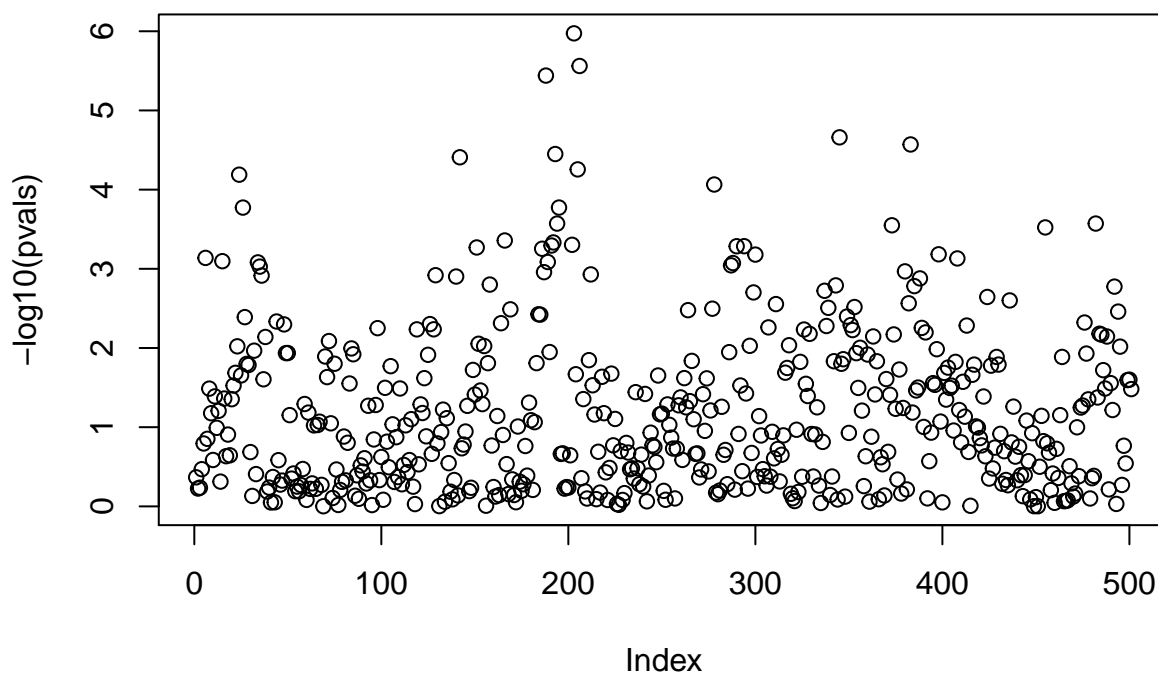
##
##      F      M <NA>
##    44     73     70
```

Exemple

Voici un exemple de regression logistique sur le jeu de données fourni. Ici on modélise l'effet de 500 sondes sur le sexe des échantillons et on visualise l'effet spécifique de 3 sondes d'intérêt.

```
s = as.numeric(as.factor(e[,colnames(d), ]$sex)) - 1
pvals = apply(d[,3500:4000, ], 1, function(meth) {
  m = glm(s ~ meth, family = binomial(logit))
  m$coefficients
  summary(m)$coefficients
  pval = summary(m)$coefficients[2, 4]
})
plot(-log10(pvals), main = "Manhattan plot") # Visualisation of pvalues
```

Manhattan plot



```

probes = names(pvals)[order(pvals)][1:3] #Select 3 probes
layout(matrix(1:3, 1, byrow = TRUE), respect = TRUE)
for (probe in probes[1:3]) {
  meth = d[probe, ]
  plot(meth, s, main = paste0("s-meth ", probe), col = s + 1)
  m = glm(s ~ meth, family = binomial(logit))
  logitinv = function(x) 1/(1 + exp(-x))
  x = sort(meth)
  lines(x, logitinv(m$coefficients[[1]] + m$coefficients[[2]] * x), col = 2,
        lwd = 2)
  pylx = function(t, m) {
    x = m$coefficients[[1]] + m$coefficients[[2]] * t
    1/(1 + exp(-x))
  }
  suppressWarnings(arrows(meth, s, meth, pylx(meth, m), col = adjustcolor(4,
    alpha.f = 0.2), length = 0.05, lwd = 3))
  legend("bottomright", "P(Y|X)", col = 4, lty = 1, cex = 0.6)
}

```

