Age-related gene expression variations in human monocytes and T cells

140.688.01 – Statistics for Genomics

Martin Skarzynski

2018-05-17

Introduction

The MESA Epigenomics and Transcriptomics Study has been launched to investigate potential gene expression regulatory methylation sites in humans by examining the association between CpG methylation and gene expression in purified human monocytes and T cells from a large study population (community-dwelling participants in the Multi-Ethnic Study of Atherosclerosis (MESA)). The MESA Epigenomics and Transcriptomics Study was funded by a National Heart, Lung and Blood Institute grant (R01HL101250) through the NIH Roadmap Epigenomics Program in 2009.

The publicly available data from the original study include transcriptomic and methylomic data from CD4+ samples, collected from 214 individuals. Peripheral T cells were isolated from blood (Exam 5) with anti-CD4 coated magnetic beads, and the Illumina HumanHT-12 v4 Expression BeadChip and the Illumina HumanMethylation450 BeadChip were used to provide genome-wide coverage of mRNA expression and DNA methylation, respectively.

Data

These data enabled the identification of CpG loci associated with age (age-DMR) and age-associated and expression-associated methylation sites (age-eMS), whose degree of methylation was associated with age and cis-gene expression (+/- 1Mb), after adjusting for other covariates such as race, gender, study site, and sample contaimination with B-cells, monocytes, natural killer cells, and neutrophils. To estimate residual sample contamination for monocyte data analysis, separate enrichment scores for neutrophils, B cells, T cells, and natural killer cells were included. The participant race, gender, and study site were provided as 'raceGenderSite variable', which represents a combination of those factors. A detailed description of each sample characteristics is included in the 'README.txt'.

Packages

Processing

- `beadarray` R package
  - QC analyses and bead-type summarization (average bead signal for each type after outlier removal)
- `limma` R package - normal-exponential convolution model analysis to estimate non-negative signal,
  - quantile normalization using all probes (gene and control, detected and not detected) and samples,
  - addition of a recommended (small) offset,
  - log2 transformation and elimination of control probe data from the normalized expression matrix.
- `lumi` package
  - Illumina HumanMethylation450 BeadChip technology employs a two-channel system and uses both Infinium I and II assays; normalization was performed.
- `MassiR` package

- to create a sex variable (a composite variable of race, gender and study site was included in the original dataset)

Results

Conclusions

We were able to predict sex using the `MassiR` package. There is difference in gene expression by age and sex in monocytes, but not in T cells. After turning age into a binary variable by the median, fewer differentially expressed genes are observed.

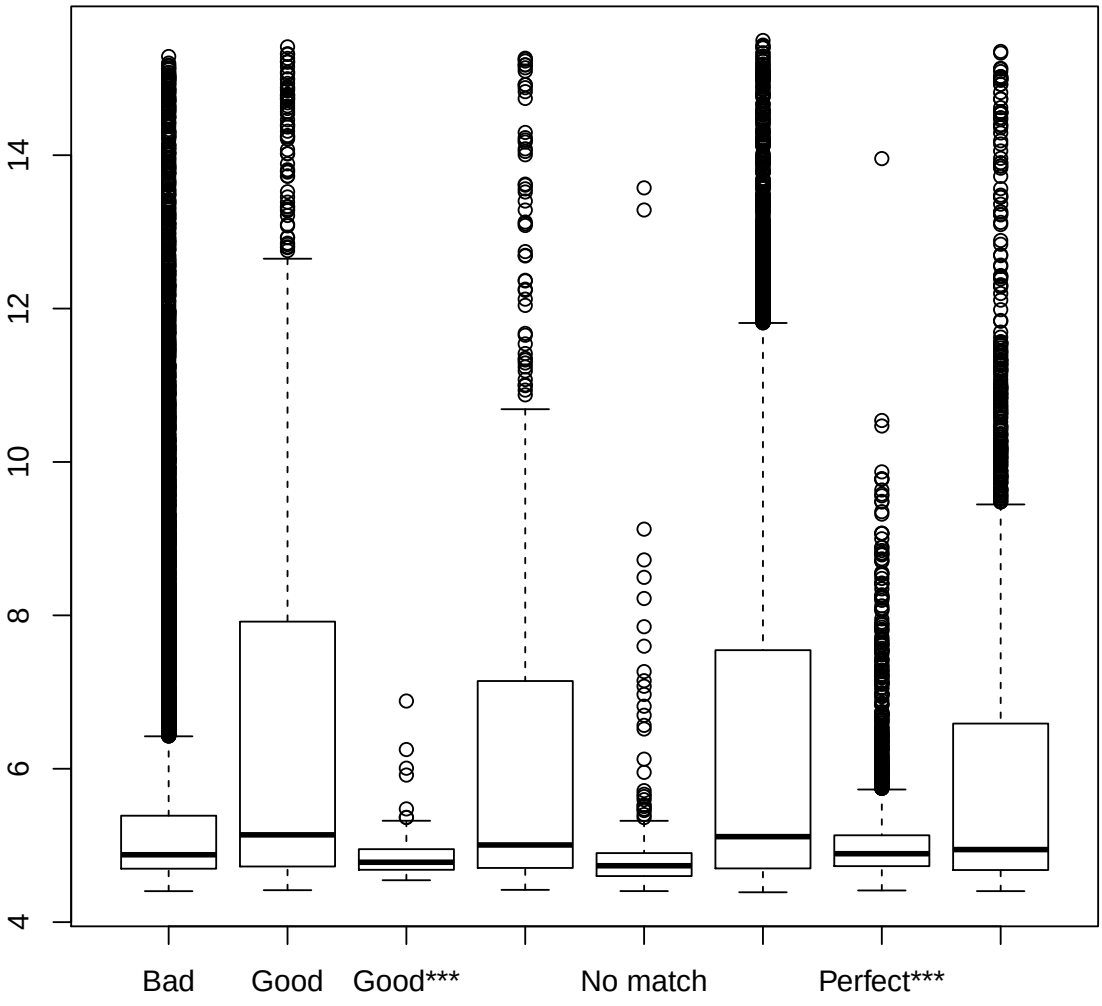

Figure 1: Microarray probe qualities

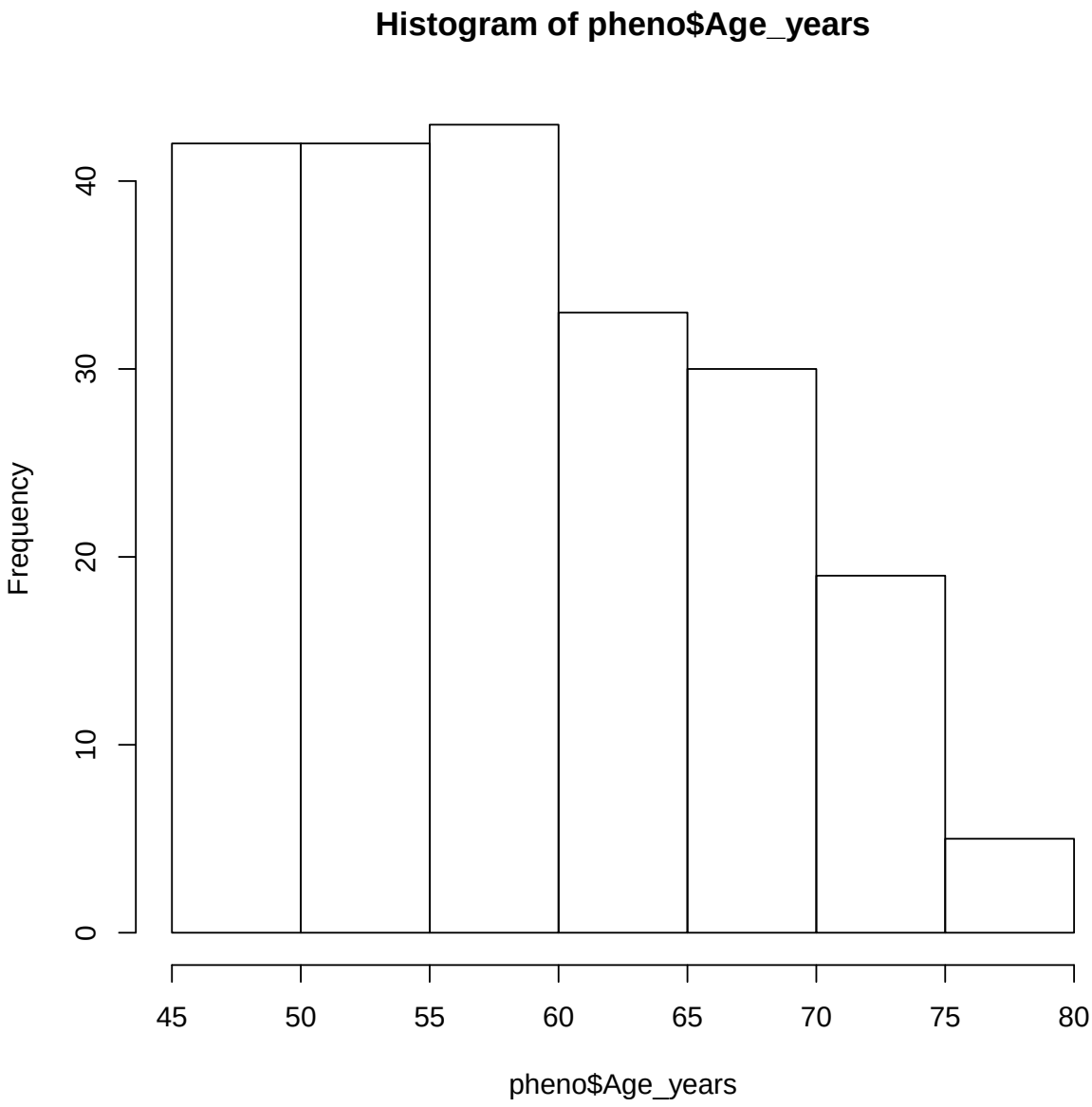

Figure 2: Age distribution of study participants

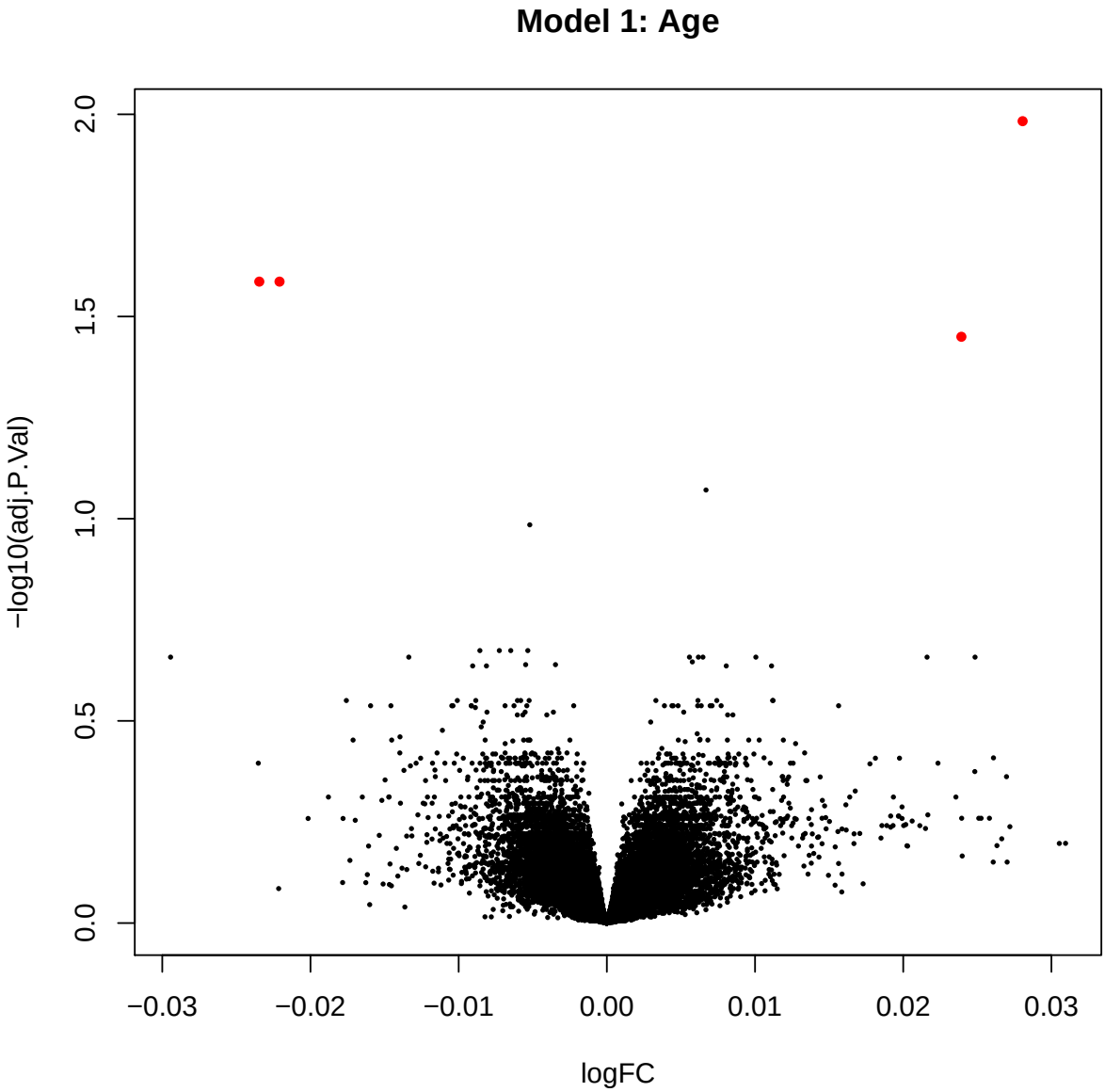

Figure 3: Volcano plot of Model 1 (Continuous Age) applied to T cells

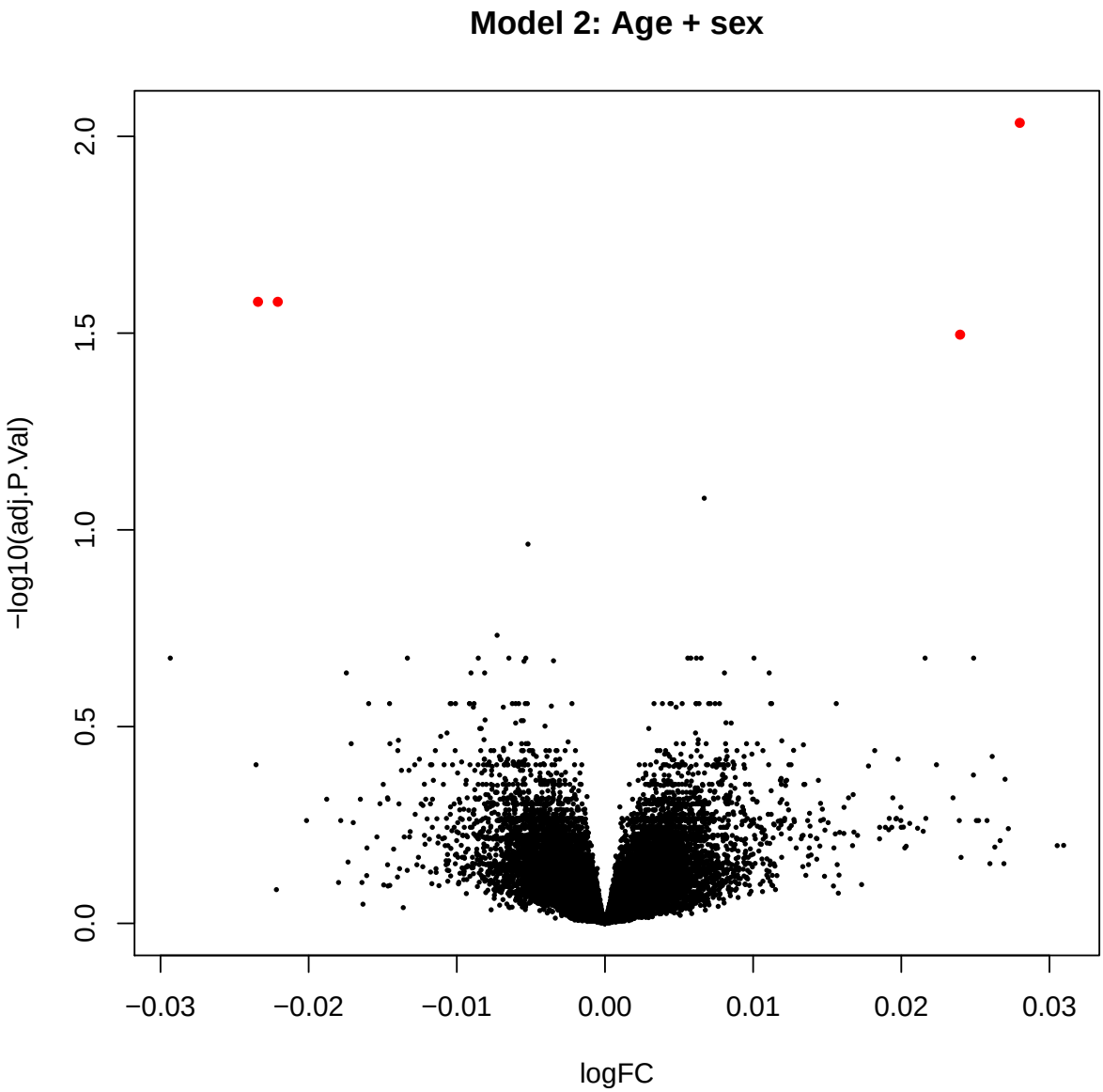

Figure 4: Volcano plot of Model 2 (Continuous Age + Sex) applied to T cells

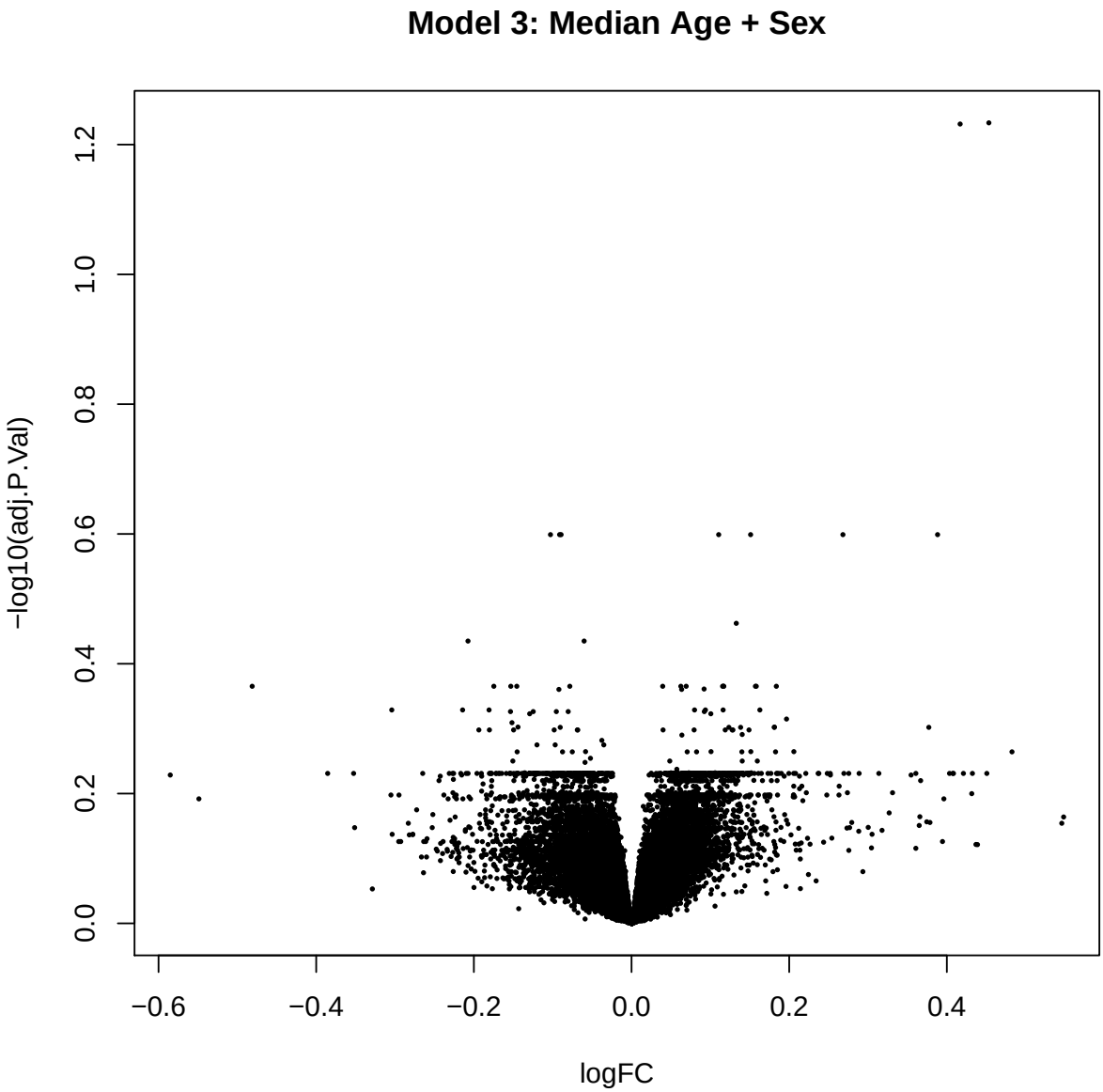

Figure 5: Volcano plot of Model 3 (Categorical Age + Sex) applied to T cells

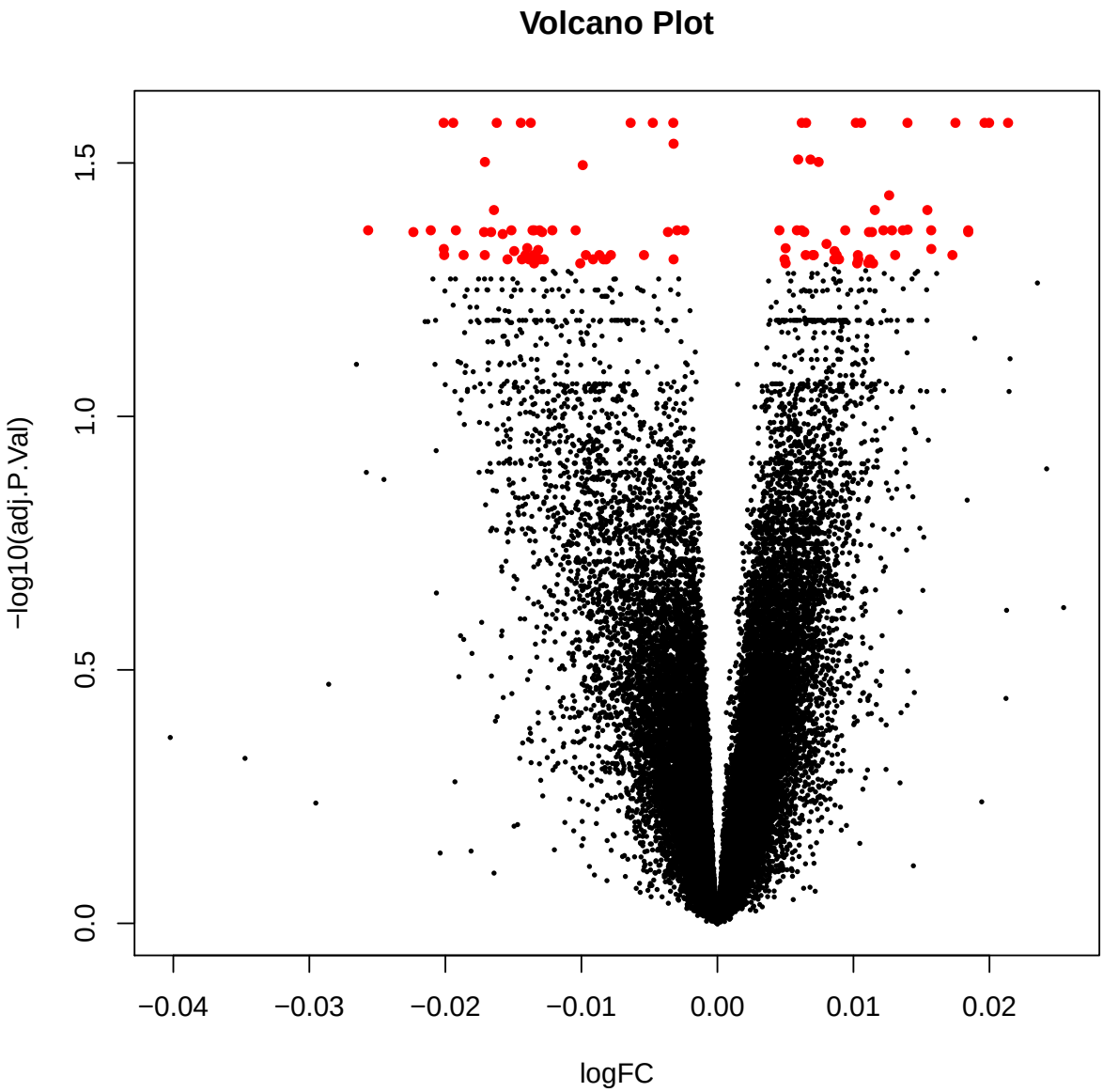

Figure 6: Volcano plot of Model 1 (Continuous Age) applied to monocytes

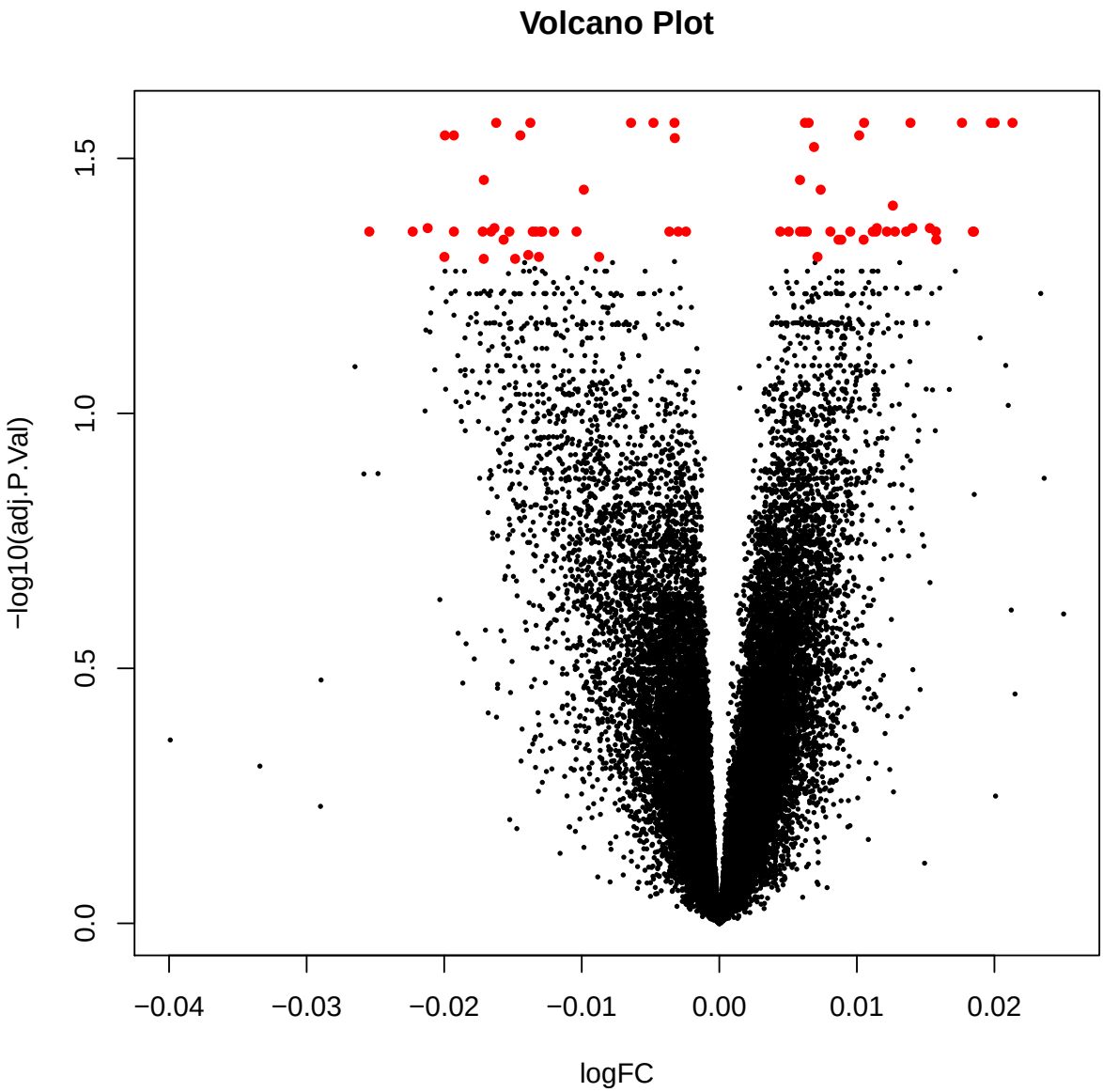

Figure 7: Volcano plot of Model 2 (Continuous Age + Sex) applied to monocytes

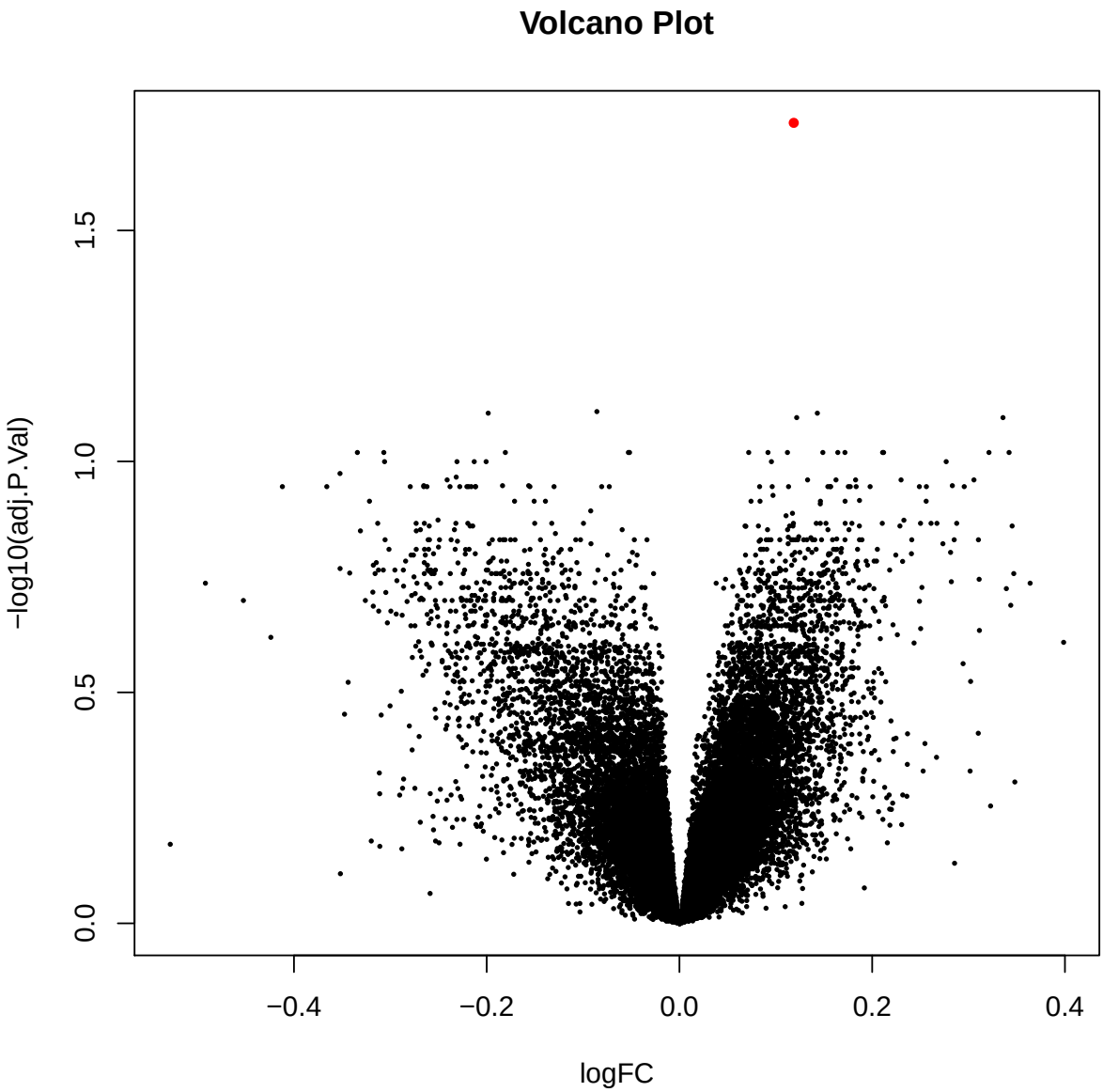

Figure 8: Volcano plot of Model 3 (Categorical Age + Sex) applied to monocytes