---

title: "DMtest"

author: "James Dai and Xiaoyu Wang"

date: "`r format(Sys.time(), '%B %d, %Y')`"

output: rmarkdown::html\_vignette

vignette: >

%\VignetteIndexEntry{DMtest}

%\VignetteEngine{knitr::knitr}

%\VignetteEncoding{UTF-8}

---

```{r, include = FALSE}

knitr::opts\_chunk$set(

collapse = TRUE,

comment = "#>"

)

```

# Introduction

Cancer development is ubiquitously associated with aberrant DNA methylation.

Noninvasive biomarkers incorporating DNA methylation signatures are of rising

interest for cancer early detection. Statistical tests for discovering differential

methylation markers have largely focused on assessing differences of mean

methylation levels, e.g., between cancer and normal samples. Cancer is a heterogeneous disease. Increased stochastic variability of cancer DNA methylation

has been observed across cancers (Hansen and others, 2011; Phipson and Oshlack, 2014), which

may reflect adaptation to local tumor environments in the carcinogenesis process. To date, differentially

variable CpG (DVC) and excessive outliers have been examined in tumor-adjacent normal tissue samples and in cancer precursors (Teschendor\_ and Widschwendter,

2012; Teschendor\_ and others, 2016), with the potential of identifying early detection

markers for the risk of progression to cancer.

In Dai et al (2021), we propose a joint constrained hypothesis test for hypermethylation and hypervariable CpG methylation (DMVC+) cites in a high-throughput profiling experiment. In the DMtest R package we implemented the constrained hypothesis test, along with the standard tests for

DMC and DVC. We also implemented another constrained test where there is no constraint for mean difference, only increased variability.

As shown in Dai et al (2021), the proposed joint tests substantially improved detection power in

simulation studies and the TCGA data example, yielding more cancer CpG markers than the standard DMC and DVC tests.

# Example

The following example takes the DNA methylation data from 334 samples of TCGA colorectal cancer samples (TCGA-COAD);

In the illustration we use representative 1000 CpG probes to save time. For genome-wide data with potentially $>$ 500,000 CpGs,

users can invoke parallel computing mode by setting appropriate numbers of cores.

```{r setup}

library(DMtest)

#load example data

data(beta)

dim(beta)

data("covariate")

dim(covariate)

#compute p-values

out=dmvc(beta=beta,covariate=covariate)

head(out)

```

# Reference

James Y. Dai et al. Incorporating increased variability in discovering

cancer methylation markers,

Biostatistics 2021, submitted.