



## Comportamento dos estimadores de metilação e hidroximetilação de DNA na presença de correlação.

Nicole N. Silva\*, Samara F. Kiihl.

### Resumo

Recentemente, além dos estudos de perfis de metilação de DNA na associação de câncer e demais doenças, tem sido crescente o interesse em perfis de hidroximetilação de DNA. Diferentes métodos foram desenvolvidos para permitir que metilação e hidroximetilação possam ser medidos simultaneamente, mas a metodologia estatística para resumir esses dados e produzir estimadores consistentes para a proporção de metilação/hidroximetilação em um dado sítio ainda não se consolidou. Dessa forma, o objetivo principal deste trabalho é estudar o comportamento dos estimadores das proporções de 5-mC e 5-hmC na presença de correlação nos dados provenientes dos diferentes métodos de conversão de DNA frequentemente utilizados (BS, oxBS e TAB).

### Palavras-chave:

Metilação de DNA, hidroximetilação de DNA, máxima verossimilhança.

### Introdução

A metilação e a hidroximetilação do DNA são mecanismos epigenéticos que possuem um papel fundamental na expressão gênica e diferenciação celular (Mariani et al. 2013; Tahiliani et al. 2009). Perfis de metilação e hidroximetilação conjuntos podem ser obtidos combinando os métodos convencionais (BS) com os métodos modificados (oxBS ou TAB). Entretanto, estas combinações produzem estimadores inconsistentes para calcular as proporções de 5-mC (citosinas metiladas) e 5-hmC (citosinas hidroximetiladas) (Qu et al. 2013). Desta forma, alguns algoritmos foram desenvolvidos recentemente para estimar 5-mC e 5-hmC de maneira consistente. É importante ressaltar que todos esses algoritmos assumem que as medidas obtidas de cada método de conversão são independentes. No entanto, quando um estudo de metilação e hidroximetilação é realizado utilizando uma combinação de dois métodos, a mesma amostra de DNA é dividida, sendo cada parte convertida por meio de um método diferente e as medidas obtidas podem apresentar correlação já que são provenientes da mesma amostra de DNA.

Assim, o objetivo principal deste trabalho é estudar o comportamento dos estimadores das proporções de 5-mC e 5-hmC na presença de correlação nos dados provenientes dos diferentes métodos de conversão de DNA utilizados (BS, oxBS e TAB). A primeira parte do projeto consiste em analisar dados reais disponibilizados publicamente para quantificar a correlação entre as medidas dos diferentes métodos de conversão do DNA. Em seguida, realizar um estudo com simulação para avaliar o impacto da correlação nos estimadores obtidos pelos algoritmos MLML e *naive* disponíveis na linguagem R.

### Resultados e Discussão

Quando as réplicas da amostra de DNA são tratadas com o reagente bissulfito (BS), as intensidades de metilação obtidas correspondem a soma de 5-mC com 5-hmC, o que não permite a discriminação entre as regiões

metiladas e as regiões hidroximetiladas. Já quando as amostras são submetidas ao tratamento de OxBS, as intensidades de metilação referem-se apenas a proporção de 5-mC enquanto as amostras tratadas com pelo método TAB revelam os índices de 5-hmC apenas. Perfis de metilação e hidroximetilação conjuntos podem ser obtidos combinando os métodos convencionais (BS) com os métodos modificados (oxBS ou TAB). Conforme a suposição no início da pesquisa, os dados apresentaram altos níveis de correlação nas amostras tratadas com BS e OxBS (cerca de 0,75), enquanto as amostras tratadas com TAB e BS apresentaram índices em torno de 0,40. De acordo com o estudo de simulação realizado, pode-se observar que o comportamento dos estimadores utilizando os algoritmos *naive* e MLML são semelhantes ao comportamento desses estimadores quando a correlação não é considerada.

### Conclusões

Conclui-se, portanto, que, embora as amostras utilizadas no estudo apresentem correlação, não há evidências para afirmar que a correlação altera o comportamento dos estimadores obtidos pelos métodos MLML e *naive* já que ambos algoritmos apresentam resultados semelhantes aos resultados que não consideraram a correlação.

### Agradecimentos

Ao CNPq, órgão que financiou a bolsa de estudos por meio do Programa Institucional de Bolsas de Iniciação Científica (PIBIC) e à professora Samara F. Kiihl que me orientou durante a realização deste trabalho.

KIIHL, Samara F. et al. MLML2R: an R package for maximum likelihood estimation of DNA methylation and hydroxymethylation proportions. *Statistical applications in genetics and molecular biology*, v. 18, n. 1, 2019.

MARIANI, Christopher et al. Alterations of 5-hydroxymethylcytosine in human cancers. *Cancers*, v. 5, n. 3, p. 786-814, 2013.

QU, Jianghan et al. MLML: consistent simultaneous estimates of DNA methylation and hydroxymethylation. *Bioinformatics*, v. 29, n. 20, p. 2645-2646, 2013

TAHILIANI, Mamta et al. Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science*, v. 324, n. 5929, p. 930-935, 2009.