Supplementary material

**Navigating the hydroxymethylome: some experimental biases and quality control tools for the tandem OxBS Illumina microarrays**

Contents

[Supplementary Table 1 Multivariable linear regression model outputs 3](#_Toc80337763)

[Supplementary Table 2 Linear mixed effect model outputs 4](#_Toc80337764)

[Supplementary Table 3. Pairwise Wilcoxon rank sum test p-values from comparing the internal normalization control probe (NORM\_A/G/C/T) intensities in BS converted and oxBS converted ccRCC samples. 5](#_Toc80337765)

[Supplementary Figure 1. The probe call rate difference across all data sets 6](#_Toc80337766)

[Supplementary Figure 2. Principal component regression analysis of technical variables in ccRCC samples 7](#_Toc80337767)

[Supplementary Figure 3. Coefficient of variation and ICC in replicate human fetal oligodendrocyte samples. 8](#_Toc80337768)

[Supplementary Figure 4. The recommended 5hmC cut-off batch 1 samples batch 2 samples using *oxBSQC* … 9](#_Toc80337769)

[Supplementary Figure 5. The distribution of 5hmC signal in low-quality probe vs high quality probe in batch 1 samples batch 2 samples using *oxBSQC* 10](#_Toc80337769)

[Supplementary Figure 6. The CpG genomic location enrichment for low-quality probes in batch1 and batch2 using *oxBSQC* 11](#_Toc80337769)

[Supplementary Figure 7. The distribution of internal normalization control probe (NORM\_A/G/C/T) intensities in BS converted and oxBS converted ccRCC samples 12](#_Toc80337769)

# **Supplementary Table 1.** Multivariable linear regression model outputs for comparing DNA extraction kit performance between BS and oxBS treated samples, with sample source adjusted.

|  |  |  |
| --- | --- | --- |
|  | **Multivariable Linear Regression Model** | |
|  | Dependent variable: | |
|  | Fraction of Low Quality Probes | |
|  | Bisulfite Converted Samples | Oxidative Bisulfite Converted Samples |
| Kit:Tandem Nucleic Acid Extraction Kit | -0.365 | -19.464\*\*\* |
|  | (-5.687, 4.956) | (-27.038, -11.890) |
| Study:ccRCC | -0.33 | 9.283\*\* |
|  | (-5.679, 5.018) | (1.671, 16.895) |
| Study:DWB | 0.698 | 8.504\*\*\* |
|  | (-1.767, 3.163) | (4.996, 12.013) |
| Study:GSE104273 | -18.581\*\*\* | -11.926\*\*\* |
|  | (-20.886, -16.276) | (-15.206, -8.646) |
| Study:GSE105109 | -7.965\*\*\* | -8.719\*\*\* |
|  | (-9.734, -6.196) | (-11.236, -6.202) |
| Study:GSE116402 | -17.796\*\*\* | 6.403 |
|  | (-24.102, -11.491) | (-2.571, 15.377) |
| Study:GSE133062 | -19.630\*\*\* | 4.204 |
|  | (-25.344, -13.916) | (-3.928, 12.337) |
| Study:GSE63179 | -4.437\*\* | -0.608 |
|  | (-8.404, -0.471) | (-6.253, 5.038) |
| Study:GSE71328 | -15.211\*\*\* | -5.409\*\*\* |
|  | (-18.041, -12.381) | (-9.437, -1.381) |
| Study:GSE71719 | -9.190\*\*\* | -6.231\*\*\* |
|  | (-11.448, -6.932) | (-9.445, -3.017) |
| Study:GSE73895 | -6.518\*\*\* | 5.291\*\*\* |
|  | (-8.657, -4.379) | (2.247, 8.336) |
| Study:PedCNS | -0.471 | 4.289\*\*\* |
|  | (-2.508, 1.566) | (1.390, 7.187) |
| Constant | 98.192\*\*\* | 83.051\*\*\* |
|  | (96.500, 99.883) | (80.644, 85.458) |
| Observations | 417 | 417 |
| R2 | 0.728 | 0.627 |
| Adjusted R2 | 0.72 | 0.616 |
| Residual Std. Error (df = 404) | 3.661 | 5.21 |
| F Statistic (df = 12; 404) | 90.091\*\*\* | 56.671\*\*\* |
| Note: \*p<0.1; \*\*p<0.05; \*\*\*p<0.01 |  |  |

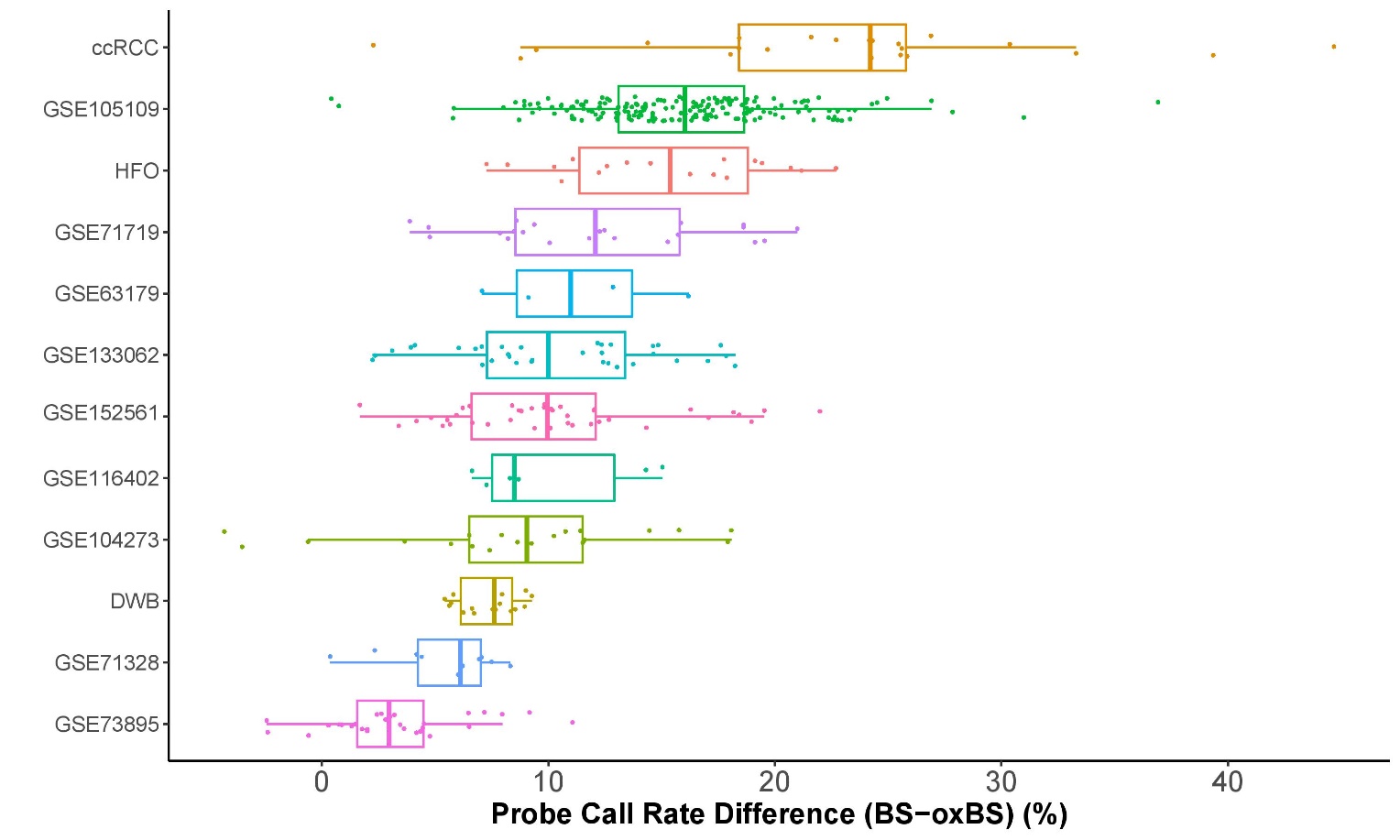
# **Supplementary Table 2.** Linear mixed effect model outputs for comparing DNA extraction kit performance between BS and oxBS treated samples, with sample source included as a random intercept.

|  |  |  |
| --- | --- | --- |
|  | **Linear Mixed Effect Model** |  |
|  | Dependent variable: | |
|  | Fraction of Low Quality Probes | |
|  | Bisulfite Converted Samples | Oxidative Bisulfite Converted Samples |
| Kit:Tandem Nucleic Acid Extraction Kit | -1.704 | -15.935\*\*\* |
|  | (-6.435, 3.026) | (-21.708, -10.162) |
| Constant | 90.240\*\*\* | 82.599\*\*\* |
|  | (85.753, 94.727) | (78.610, 86.589) |
| Observations | 417 | 417 |
| Log Likelihood | -1,155.39 | -1,296.61 |
| Akaike Inf. Crit. | 2,318.77 | 2,601.22 |
| Bayesian Inf. Crit. | 2,334.89 | 2,617.33 |
| Note: \*p<0.1; \*\*p<0.05; \*\*\*p<0.01 |  |  |

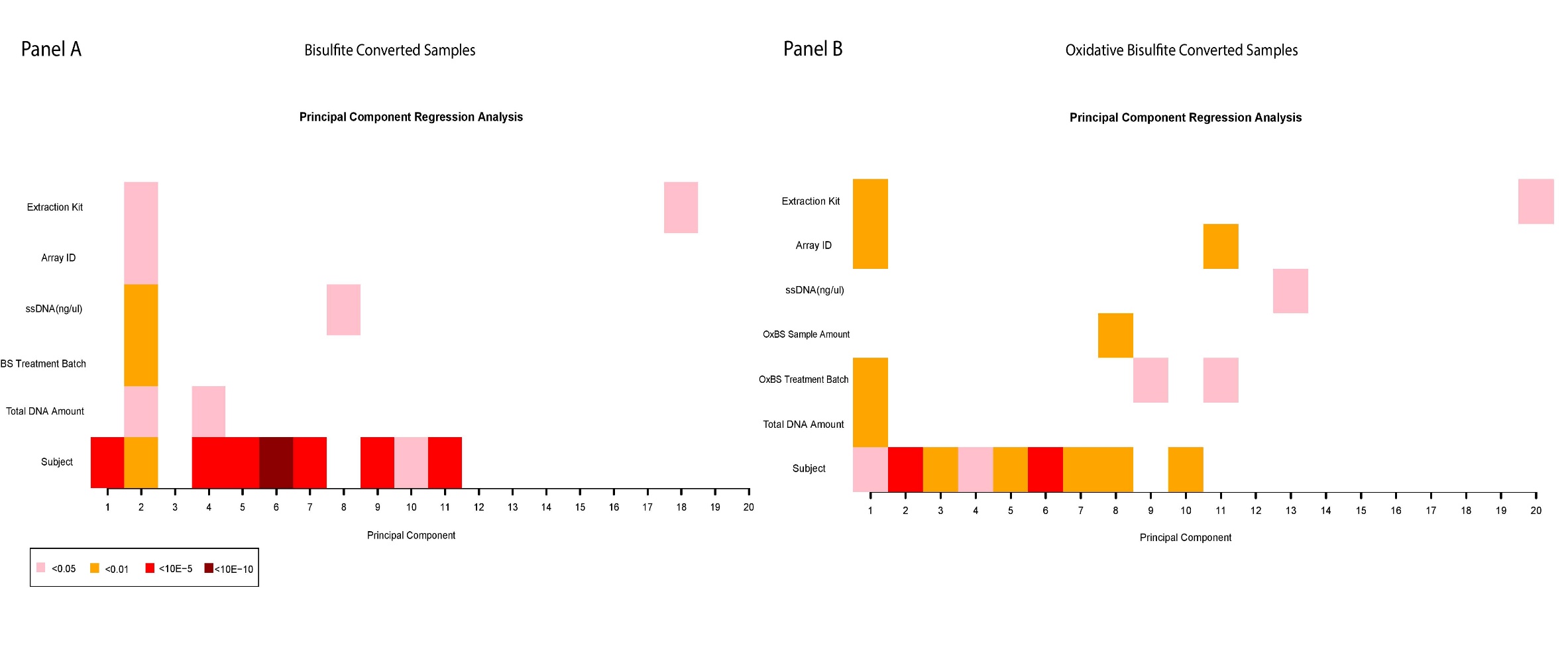
# **Supplementary Table 3.** Pairwise Wilcoxon rank sum test p-values from comparing the internal normalization control probe (NORM\_A/G/C/T) intensities in BS converted and oxBS converted ccRCC samples.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| ID | Norm\_A Red P-value | Norm\_C Green P-value | Norm\_G Green P-value | Norm\_T Red P-value |
| RCD002 | 7.45E-09 | 1.90E-11 | 7.45E-09 | 2.00E-11 |
| RCD003 | 7.45E-09 | 1.90E-11 | 7.45E-09 | 3.03E-11 |
| RCD005 | 7.45E-09 | 1.90E-11 | 7.45E-09 | 3.36E-11 |
| RCD007 | 7.45E-09 | 1.80E-11 | 7.45E-09 | 1.90E-11 |
| RCD008 | 3.73E-08 | 2.22E-11 | 1.49E-08 | 4.35E-11 |
| RCD013 | 3.73E-08 | 2.59E-11 | 2.24E-08 | 5.34E-11 |
| RCD047 | 7.45E-09 | 1.90E-11 | 7.45E-09 | 2.22E-11 |
| RCD058 | 7.45E-09 | 1.90E-11 | 7.45E-09 | 2.46E-11 |
| RCD060 | 7.45E-09 | 1.90E-11 | 7.45E-09 | 1.90E-11 |
| RCD062 | 7.45E-09 | 1.80E-11 | 7.45E-09 | 1.90E-11 |
| RCD072 | 0.012269832 | 0.003780043 | 0.052813068 | 0.000622078 |
| RCD074 | 1.42E-07 | 5.62E-11 | 4.10E-07 | 8.43E-11 |
| RCD078 | 7.45E-09 | 1.90E-11 | 7.45E-09 | 2.00E-11 |
| RCD080 | 7.45E-09 | 2.00E-11 | 7.45E-09 | 3.03E-11 |
| RRC165 | 7.45E-09 | 1.90E-11 | 7.45E-09 | 3.19E-11 |
| RRC172 | 7.45E-09 | 3.03E-11 | 7.45E-09 | 6.22E-11 |
| RRC228 | 7.45E-09 | 1.90E-11 | 7.45E-09 | 2.59E-11 |
| RRC231 | 7.45E-09 | 1.90E-11 | 7.45E-09 | 1.90E-11 |
| RRC678 | 7.45E-09 | 2.11E-11 | 7.45E-09 | 6.22E-11 |
| RRC691 | 1.49E-08 | 1.90E-11 | 2.24E-08 | 2.73E-11 |
| RRC725 | 7.45E-09 | 1.90E-11 | 7.45E-09 | 2.00E-11 |
| RRC755 | 7.45E-09 | 1.90E-11 | 7.45E-09 | 2.34E-11 |

# **Supplementary Figure 1.** The probe call rate difference across all data sets



# **Supplementary Figure 2.** Principal component regression analysis of technical variables in ccRCC samples



# **Supplementary Figure 3.** Coefficient of variation and ICC in replicate human fetal oligodendrocyte samples.

Diagram

Description automatically generated

# **Supplementary Figure 4.** The recommended 5hmC cut-off batch 1 samples batch 2 samples using *oxBSQC*.

**Chart

Description automatically generated**

# 

# **Supplementary Figure 5.** The distribution of 5hmC signal in low-quality probe vs high quality probe in batch 1 samples batch 2 samples using *oxBSQC*.

# 

# **Supplementary Figure 6.** The CpG genomic location enrichment for low-quality probes in batch1 and batch2 using *oxBSQC*.

# 

# **Supplementary Figure 7.** The distribution of internal normalization control probe (NORM\_A/G/C/T) intensities in BS converted and oxBS converted ccRCC samples.