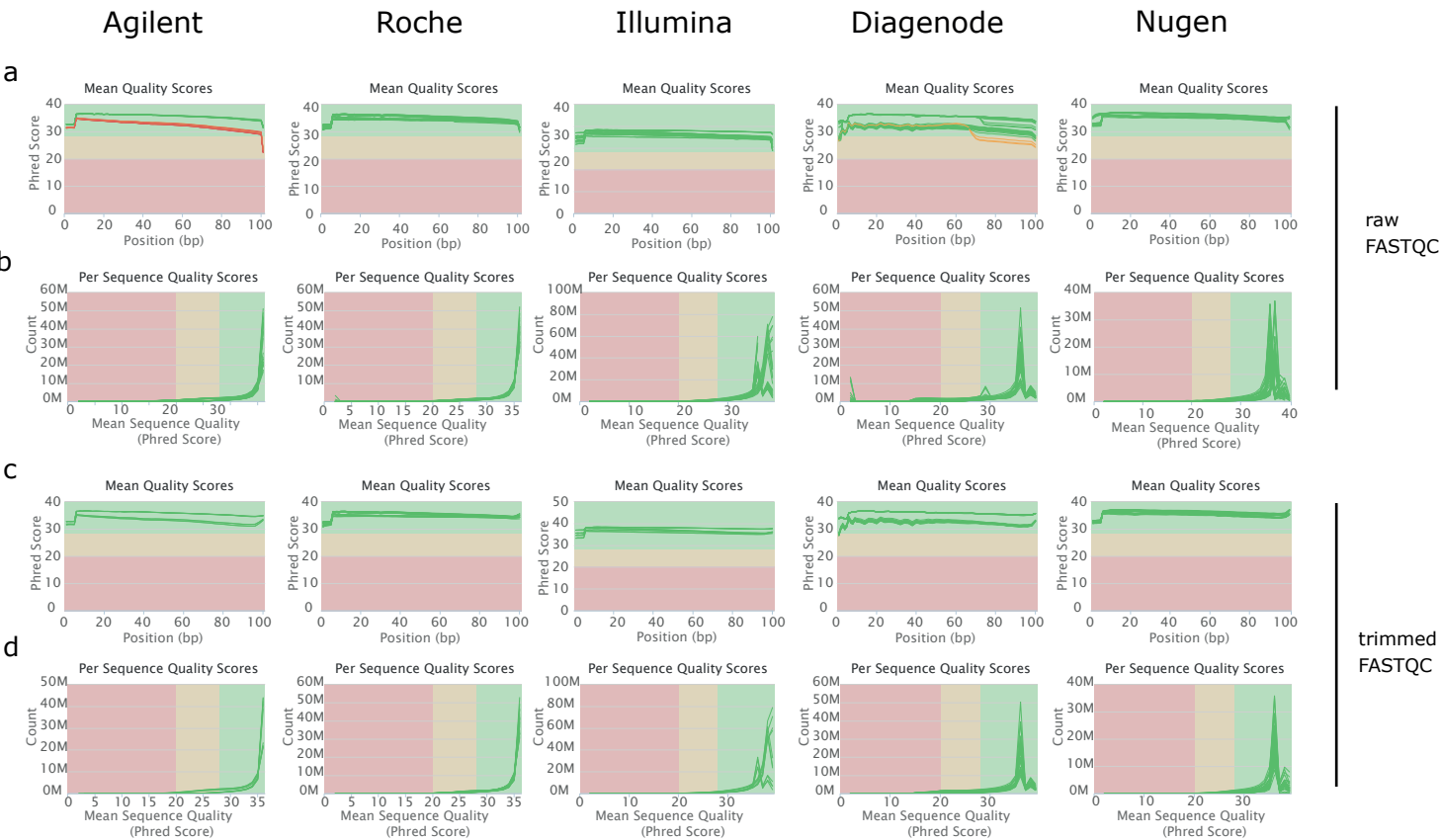

Supplementary information

**Comparison and imputation-aided
integration of five commercial platforms
for targeted DNA methylome analysis**

In the format provided by the
authors and unedited

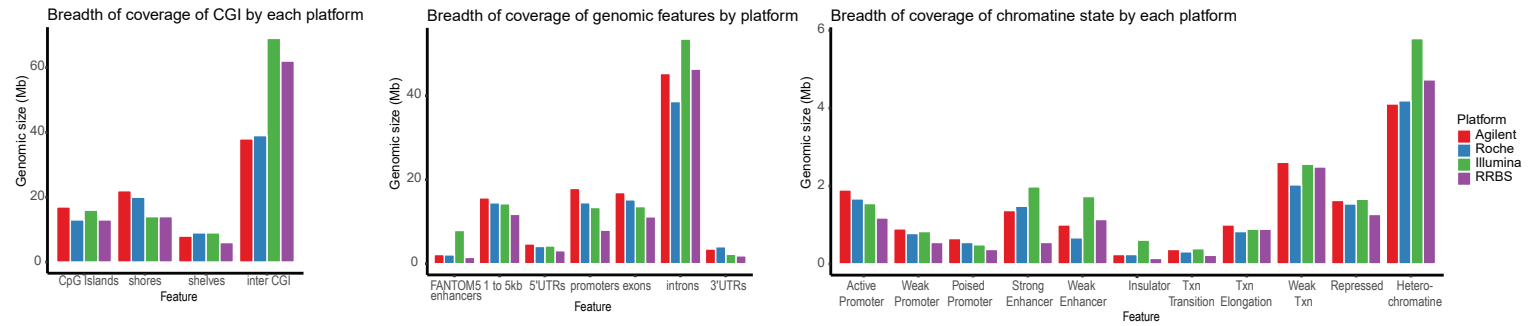
Supplementary Figure 1



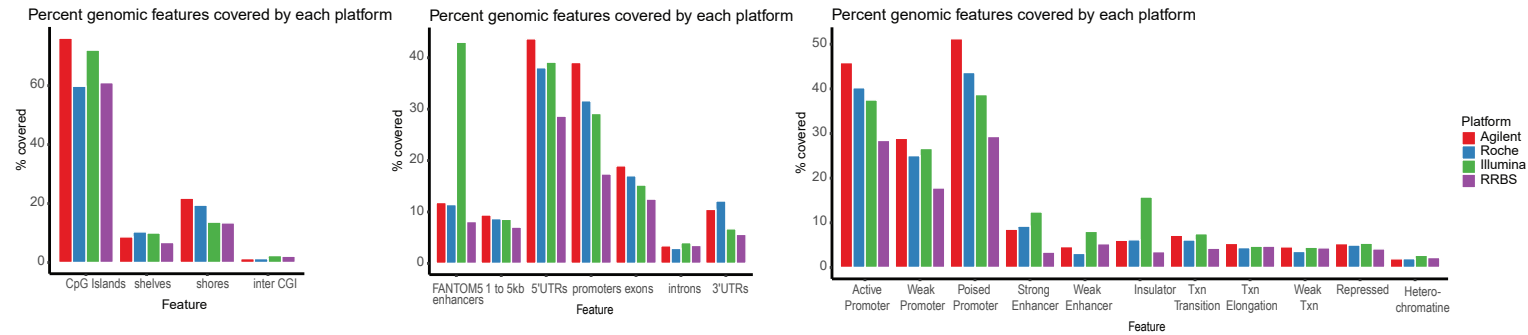
Supplementary Figure 1. Raw and trimmed sequencing data quality by the platform. **a** Sequence quality (Phred score) by read position for raw FASTQ reads by the platform; **b** The number of raw reads as a function of sequence quality (Phred score) by the platform; **c** Sequence quality (Phred score) by read position for trimmed FASTQ reads by the platform; **d** The number of trimmed reads as a function of sequence quality (Phred score) by the platform.

Supplementary Figure 2

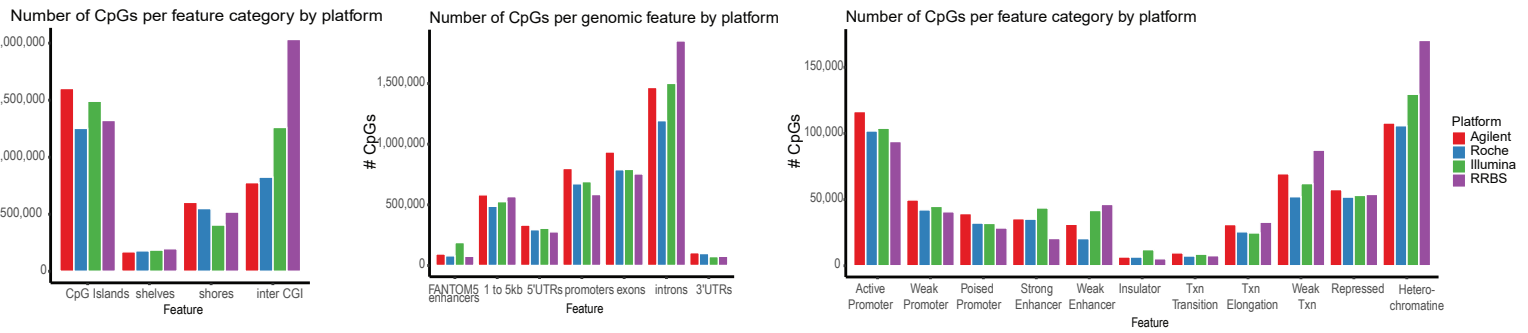
a



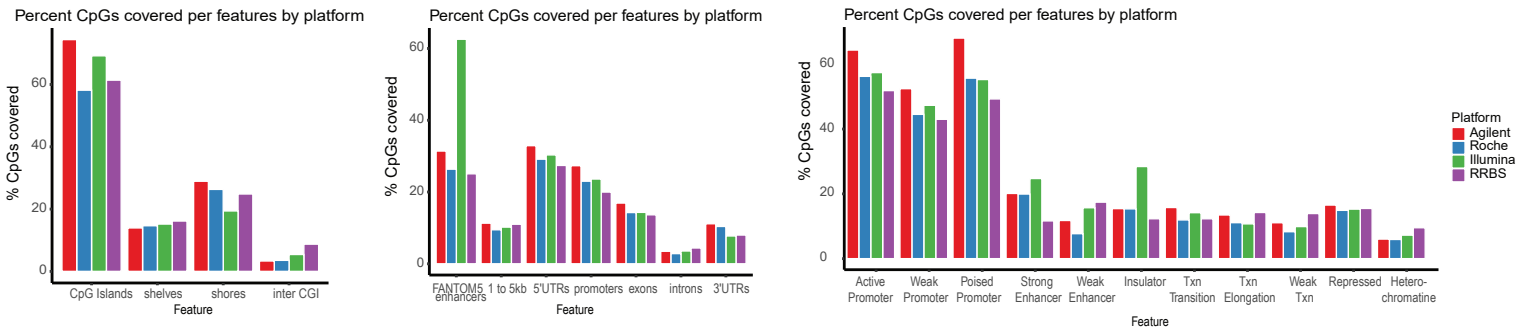
b



c

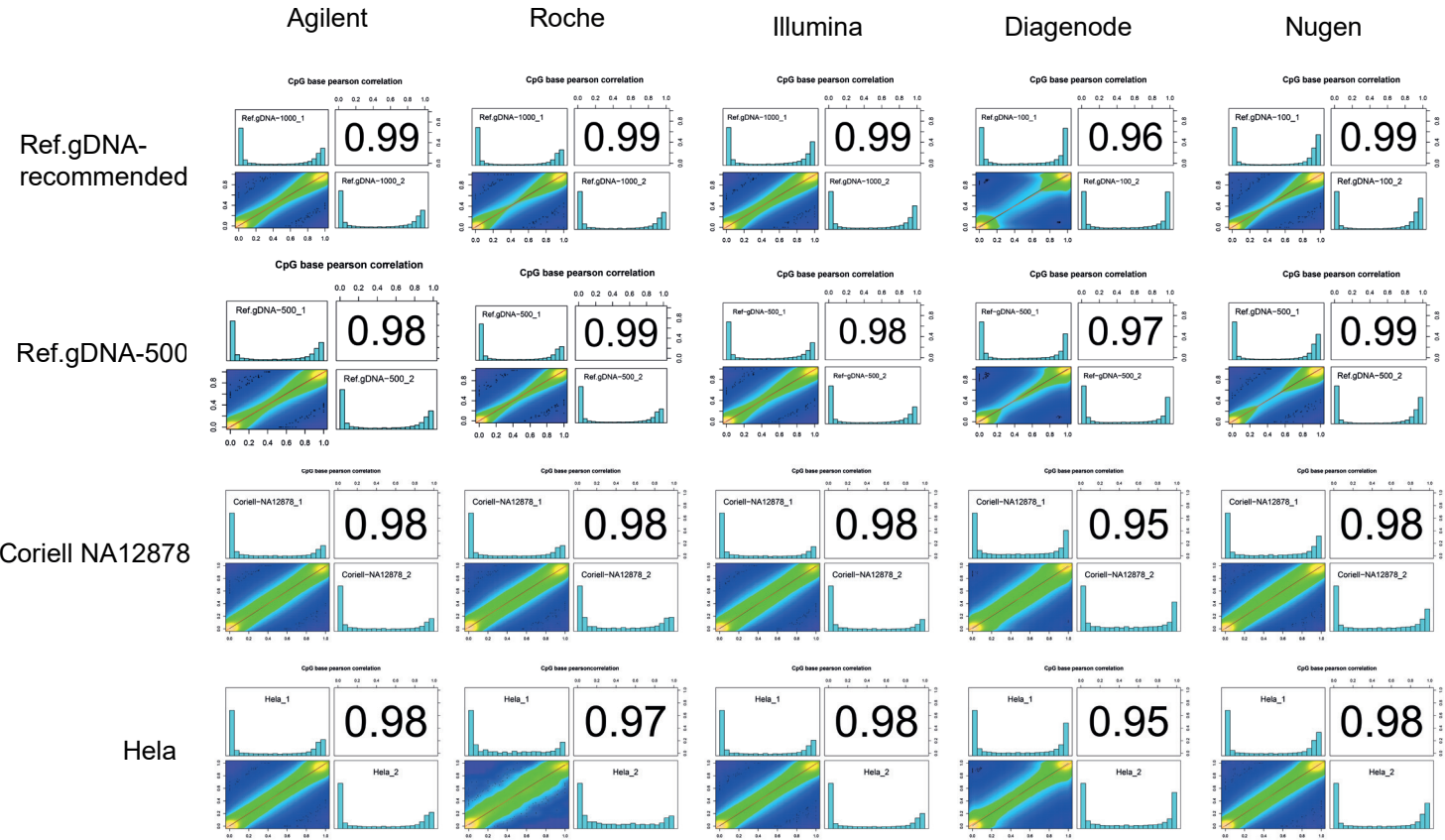


d



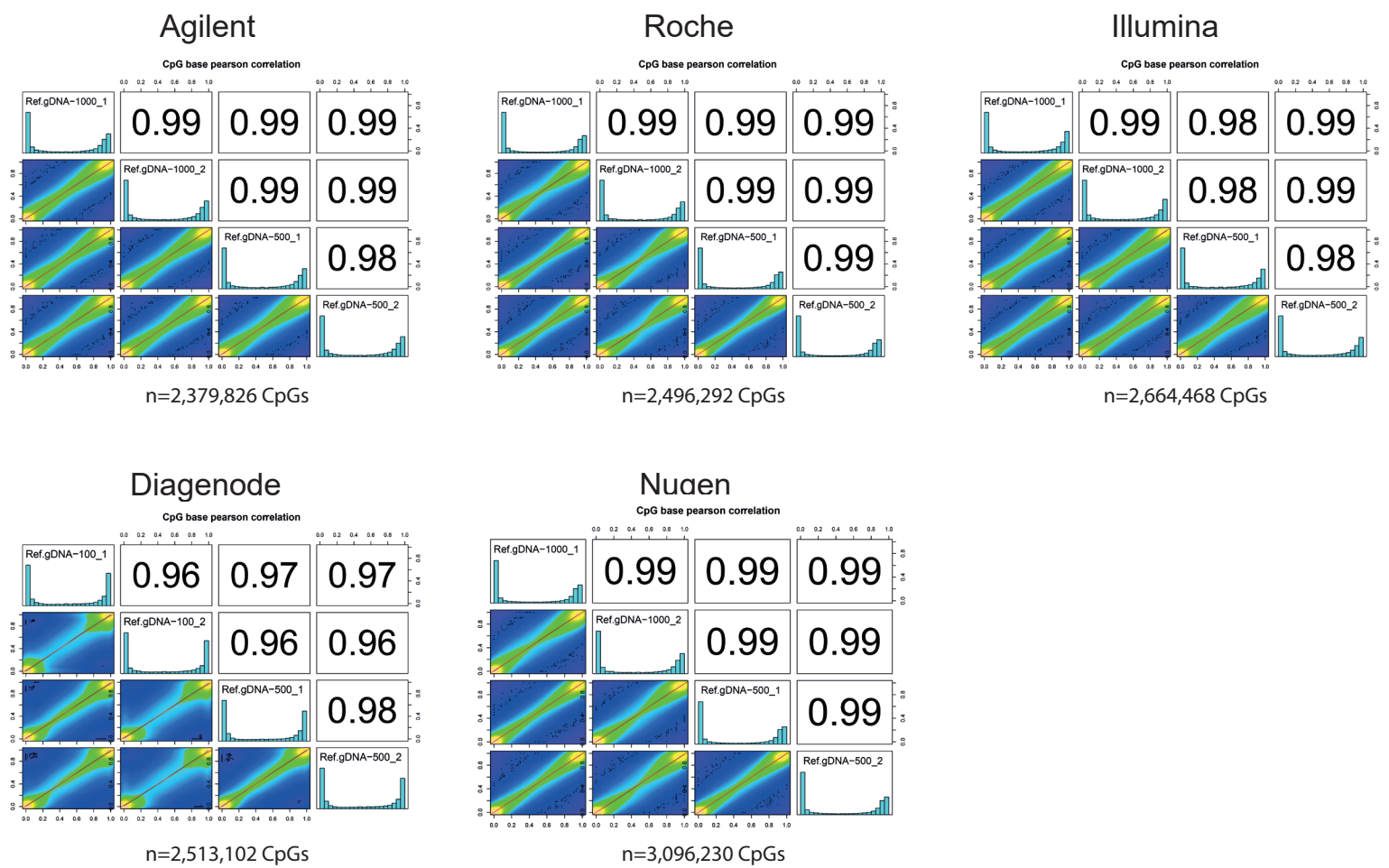
Supplementary Figure 2. Feature annotation of region and CpGs targeted by design per platform. a-b Size of the region in Mb targeted by design (a) and percent of the total size of that feature type (b) for each platform stratified by feature type CpG density (left), genic features (middle) and regulatory feature type (right) for each platform;; **c-d** Number of CpGs targeted by design (c) and percent of all CpGs in that feature type (d) stratified by feature type CpG density (left), genic features (middle) and regulatory feature type (right) for each platform

Supplementary Figure 3



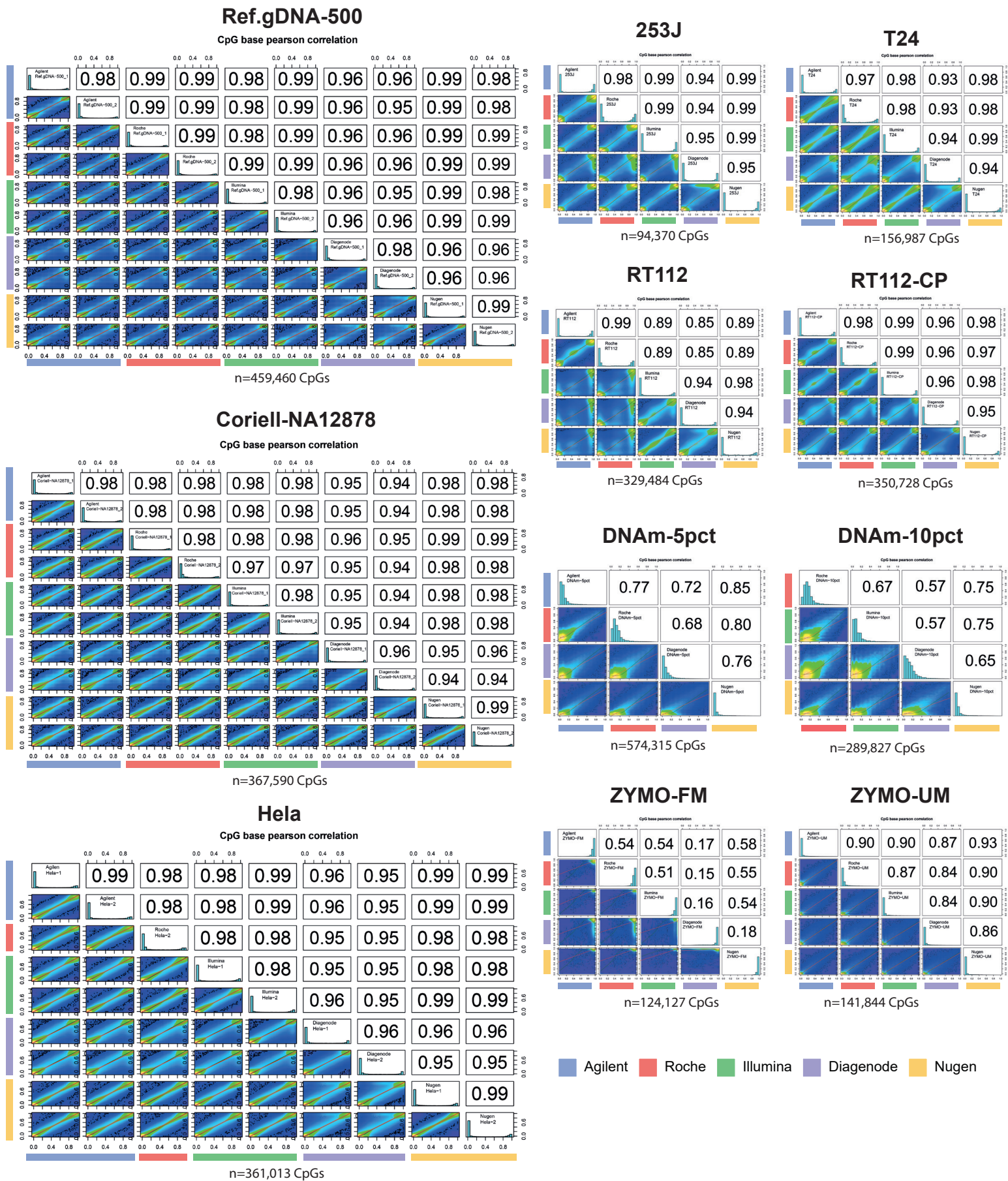
Supplementary Figure 3. Intra-platform reproducibility in DNA methylation calls for each technical replicate sample per platform. The panel shows for each sample-platform a correlogram showing pairwise Pearson correlation coefficient of the methylation levels for commonly interrogated CpGs at 10x (n); a histogram of DNA methylation beta values for each sample; and a scatterplot of DNA methylation beta values on a scale of 0 to 1.

Supplementary Figure 4



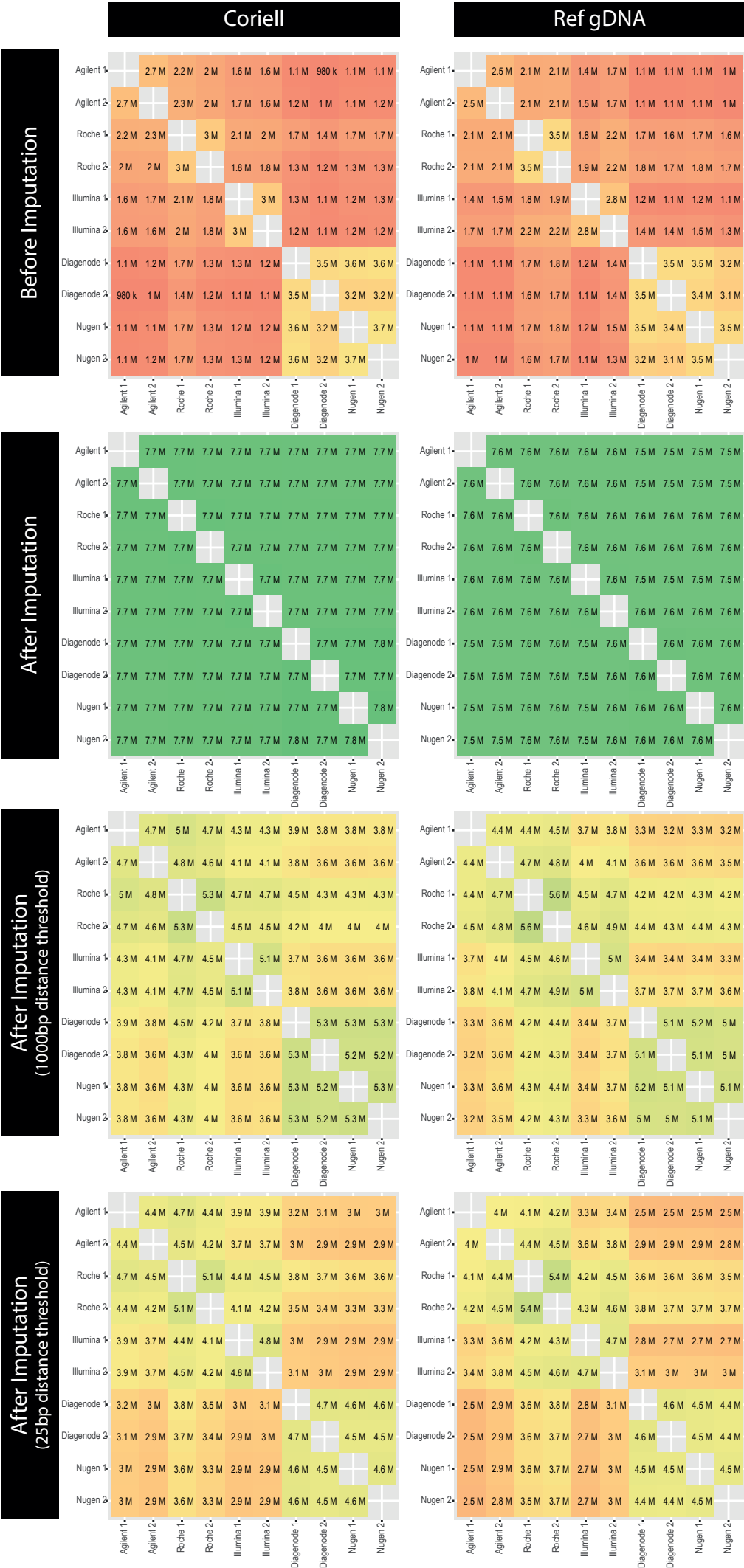
Supplementary Figure 4. The effect of sample input measured as concordance in DNA methylation levels for Ref.gDNA samples at specified inputs within each platform. The panel shows for each platform a correlogram showing pairwise Pearson correlation coefficient of the methylation levels for commonly interrogated CpGs at 10x (n); a histogram of DNA methylation beta values for each sample; and a scatterplot of DNA methylation beta values on a scale of 0 to 1.

Supplementary Figure 5

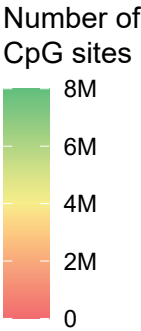


Supplementary Figure 5. Between-platform concordance in DNA methylation calls for each sample across all platforms. The panel shows for each sample-platform a correlogram showing pairwise Pearson correlation coefficient of the methylation levels for commonly interrogated CpGs at 10x (n); a histogram of DNA methylation beta values for each sample; and a scatterplot of DNA methylation beta values on a scale of 0 to 1.

Supplementary Figure 6



Supplementary Figure 6. Interoperability between platforms for Coriell NA12878 (left) and Ref.gDNA (right) before imputation (first row), after imputation without distnctce treshold (sencond row), after imputation with 1000bp distance treshold (third row) and after imputauion with 25bp distance treshold (fourth row). Correlation heatmap showing number of CpGs overlapping between pairs of sample-platforms.



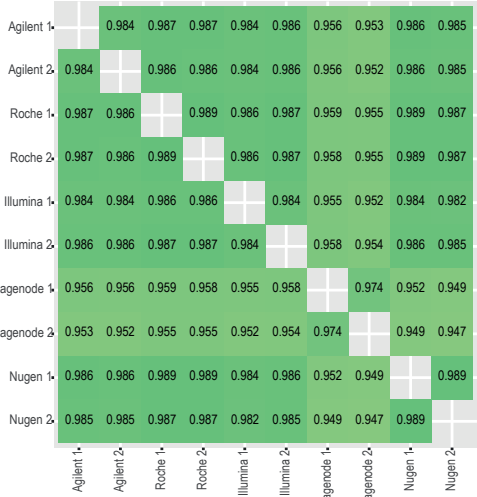
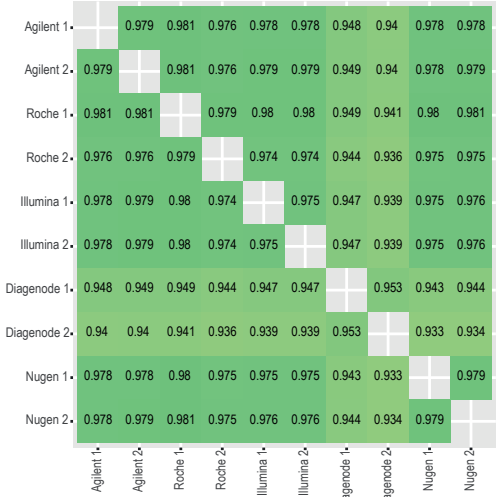
Supplementary Figure 7

Coriell

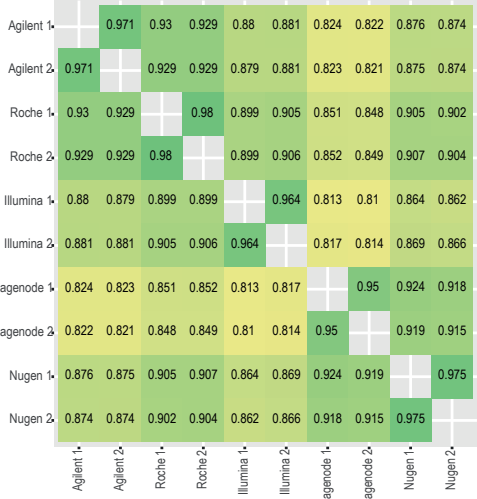
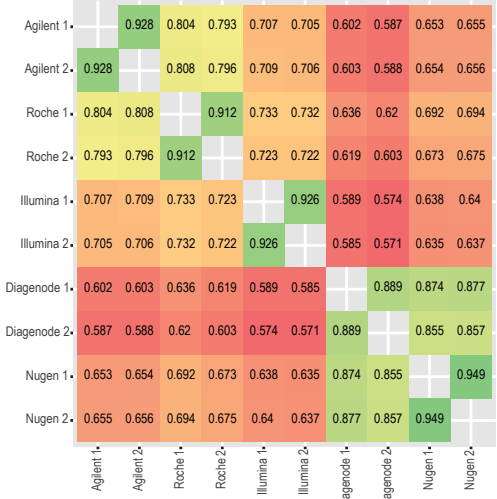
Ref gDNA

Supplementary Figure 7. Interoperability between platforms for Coriell NA12878 (left) and Ref.gDNA (right) before imputation (first row), after imputation without distnce treshold (sencond row), after imputation with 1000bp distance treshold (third row) and after imputauiion with 25bp distance treshold (fourth row). Correlation heatmap showing pairwise Pearson correlation coeafficient over CpGs common between pairs of sample-platforms.

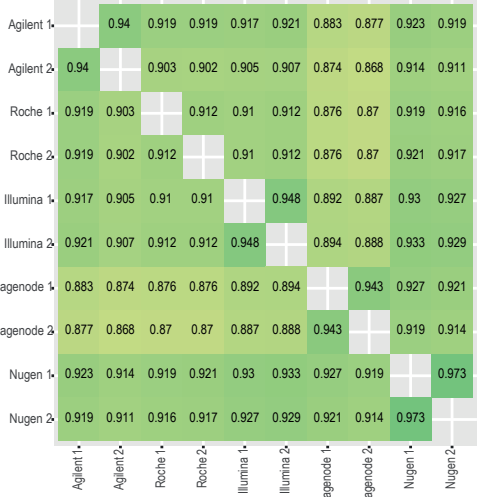
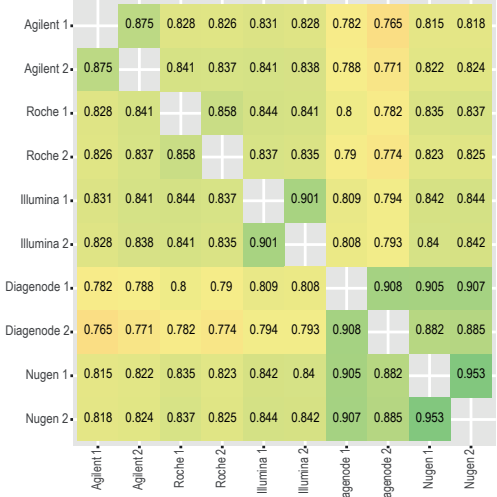
Before Imputation



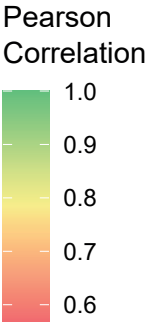
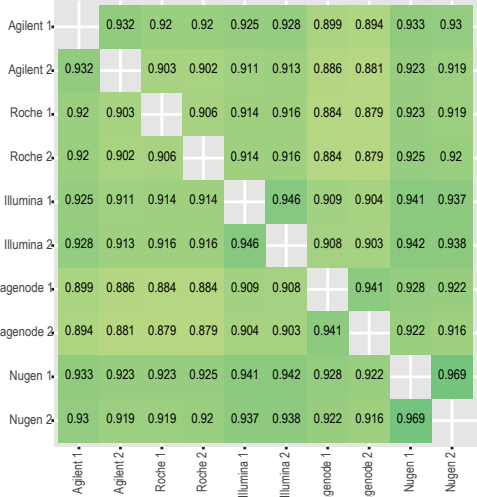
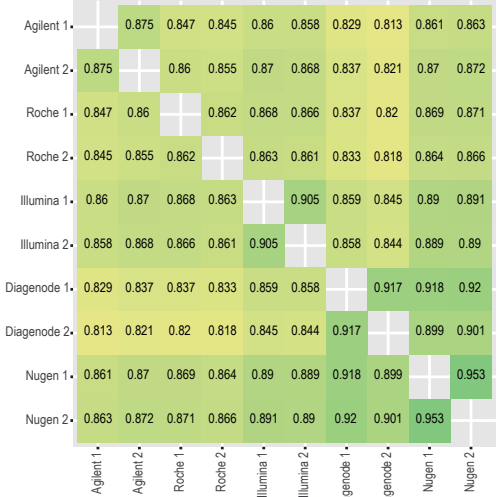
After Imputation



After Imputation
(1000bp distance threshold)



After Imputation
(25bp distance threshold)



Supplementary Figure 8

Before Imputation

After Imputation

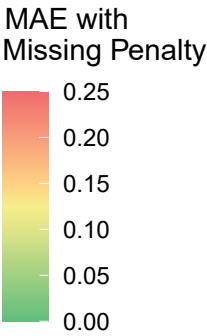
After Imputation
(1000bp distance threshold)

After Imputation
(25bp distance threshold)

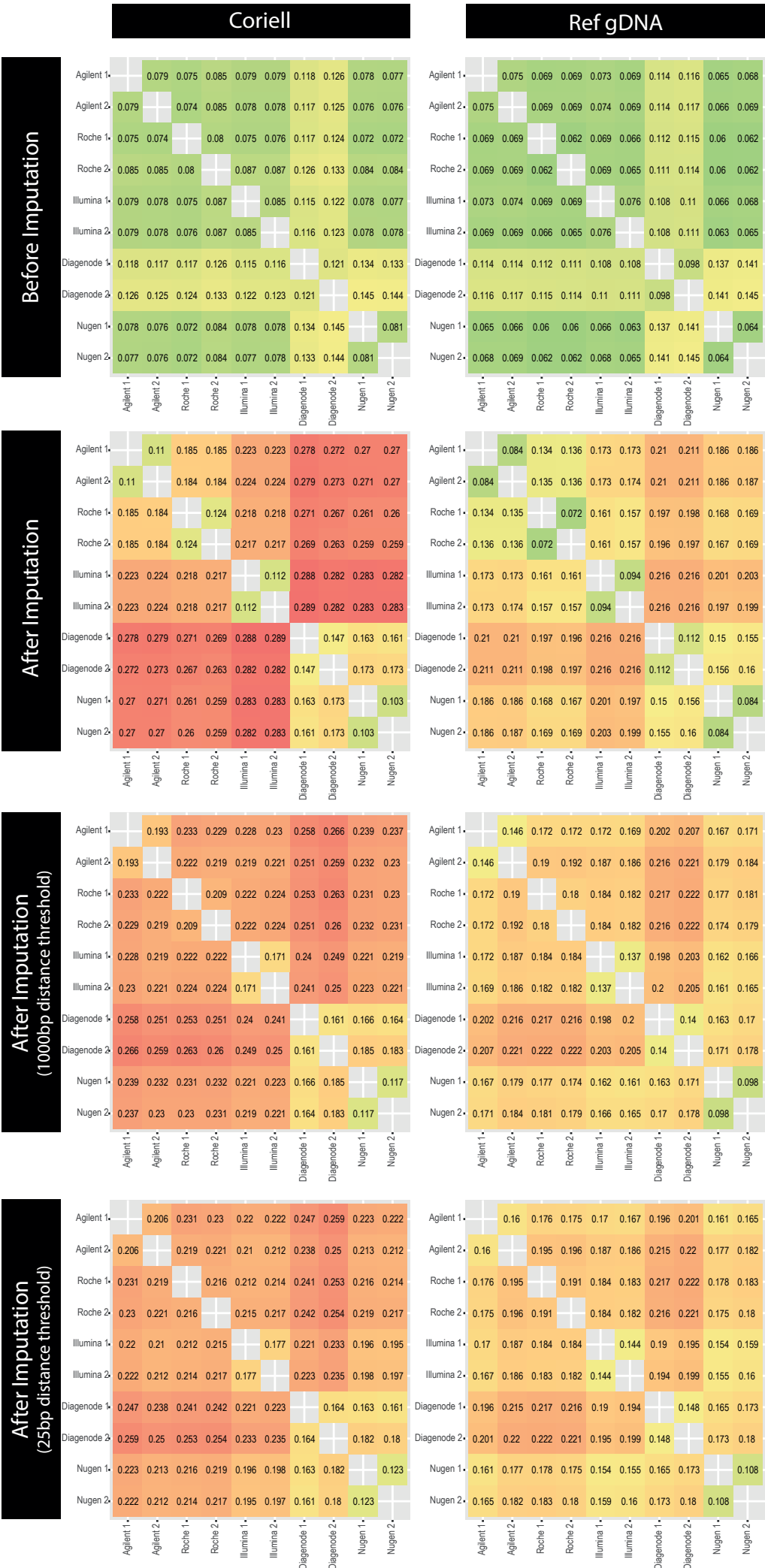
Coriell

Ref gDNA

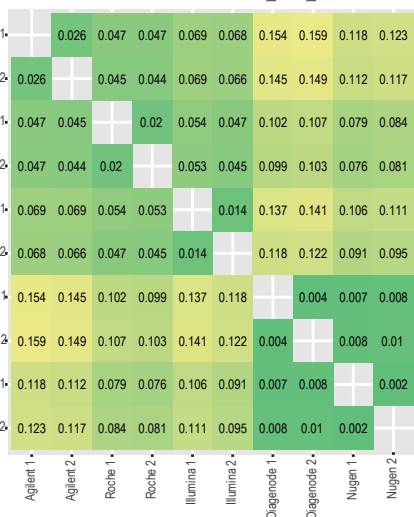
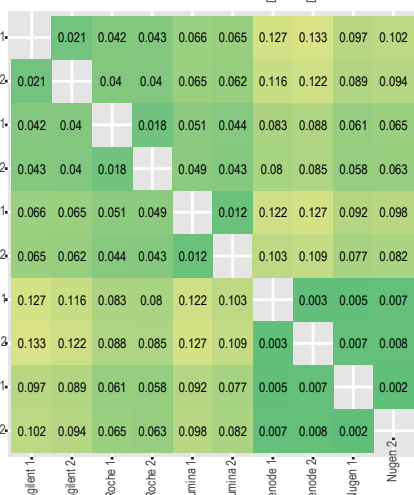
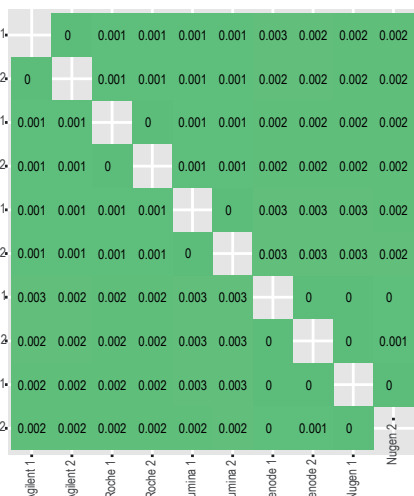
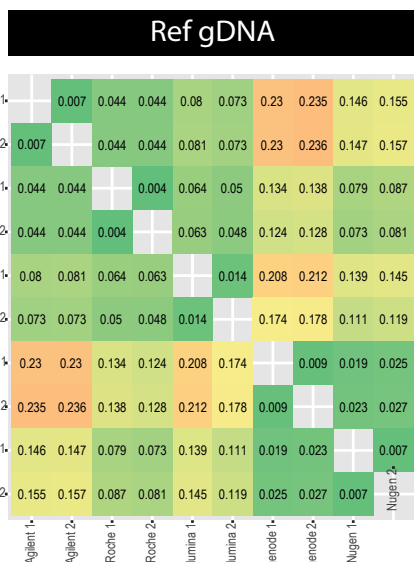
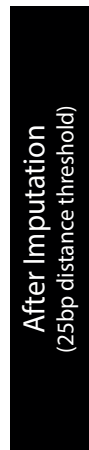
Supplementary Figure 8. Interoperability between platforms for Coriell NA12878 (left) and Ref.gDNA (right) before imputation (first row), after imputation without distnce treshold (sencond row), after imputation with 1000bp distance treshold (third row) and after imputauion with 25bp distance treshold (fourth row). Correla-tion heatmap showing pairwise mean absolute error (MAE) with missing data penalty.



Supplementary Figure 9



Before Imputation



Supplementary Figure 10. Interoperability between platforms for Coriell NA12878 (left) and Ref.gDNA (right) before imputation (first row), after imputation without distance threshold (second row), after imputation with 1000bp distance threshold (third row) and after imputation with 25bp distance threshold (fourth row). Correlation heatmap showing pairwise corrected MAE for the number of CpG sites present.