Technical Variability on the Illumina HumanMethylation450 Platform

Chris Kay[1], Ellyce Eddy[2], Ruiwei Jiang[2], Michelle Xia[3]

[1] Department of Medical Genetics, [2] Genome Science and Technology Program, [3] Department of Statistics , University of British Columbia, Vancouver , Canada

The Issue

High-throughput biology is increasingly dependent on commercial technology manufactured to produce measurements of high technical precision. Researchers need confidence that any noise observed reflects biological reality, and not the particular instrument used to assess it. Likewise, regular differences should ideally reflect real sample differences, rather than bias of a technical nature.

illumina®

Trust us. It works.

Massively parallel arrays pose a unique challenge in this regard, due to the element of chance. The more independent probes and tests performed, the more chance outliers and false discoveries result. Technical variation becomes more and more difficult to discern in high-throughput contexts, especially with novel problems and costly technologies.

In this project, we endeavored to characterize technical variation of sample replicates on the Illumina HumanMethylation450 BeadChip, introduced to the market in 2011. The HumanMethylation450 array characterizes methylation at more than 450,000 CpG sites across the human genome, more than 20 times as dense as the previous HumanMethylation27 platform. But before we collectively sprint ahead to calculate new p-values, we must ask how much all these sites vary, and how much variability exists between simple replicates. At present, each sample on the HumanMethylation450 array costs, coincidentally, $450 per run – investigators are unlikely to run samples twice unless needed.

Our Replicate Pairings

Thanks to the Kobor Lab at the Centre for Molecular Medicine and Therapeutics, we pooled a set of sample replicates for analysis of technical variation. As tabulated on the right, they are:

- **3 samples run in replicate over different batches**
- **6 samples run in replicate on different chips of the same batch**
- **6 samples run in replicate on the same chip of the same batch**

Interbatch Groups — *Replicates run a week apart*

Group1	Chip	Group2	Chip	Tissue Type
Rep01a	5998789008	Rep01b	5998789031	Buccal
Rep02b	5998789031	Rep02a	5998789029	Buccal
Rep02b	5998789031	Rep02c	6006307022	Buccal

Interchip Groups — *Replicates run on different chips in the same batch*

Group1	Chip	Group2	Chip	Tissue Type
Rep02a	5998789029	Rep02c	6006307022	Buccal
Rep05a	5808922040	Rep05b	5808922083	Buccal
Rep06c	5808922080	Rep06a	5808922076	Buccal
Rep06c	5808922080	Rep06b	5808922076	Buccal
Rep10a	6229009045	Rep10b	6229009015	Blood
Rep11a	6229009016	Rep11b	6229009015	Blood

Intrachip Groups — *Replicates run on the same chip in the same batch*

Group1	Chip	Group2	Chip	Tissue Type
Rep03a	5808922063	Rep03b	5808922063	Buccal
Rep04a	5808922054	Rep04b	5808922054	Buccal
Rep06a	5808922076	Rep06b	5808922076	Buccal
Rep07a	5808922042	Rep07b	5808922042	Buccal
Rep08a	5808922040	Rep08b	5808922040	Buccal
Rep09a	5808922041	Rep09b	5808922041	Buccal

A Note About Methylation

DNA methylation is an epigenetic modification of cytosine residues, typically at so-called CpG sites where guanine follows 5'-3'. Methylation plays a key role in developmental regulation and gene expression, and has been demonstrated to mediate environmental influences on the genome.

Replicate Correlations

To assess overall correlation of methylation between replicates, pairwise Pearson correlations were computed across all samples and replicates. Samples were clustered for comparison.

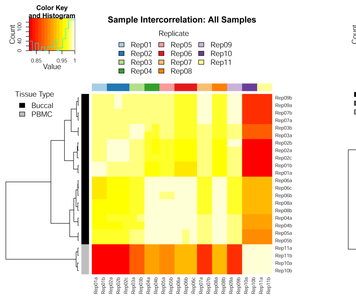


Figure 1. Correlation matrix for sample methylation between all samples and replicates. Tissue type is the strongest cluster, and replicates cluster pairwise.

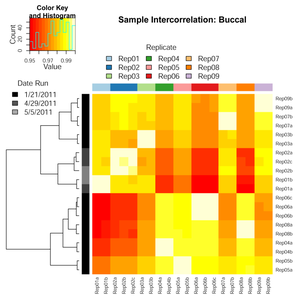


Figure 2. Correlation matrix for sample methylation with blood replicates removed. Replicates cluster together, but the strongest cluster associates with an unknown factor, likely sample gender.

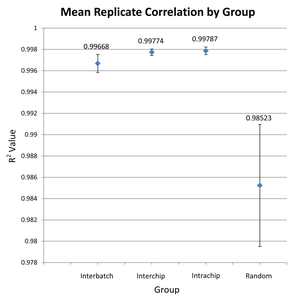


Figure 3. Mean R^2 and standard deviation by comparison group. Interbatch replicates demonstrate the greatest standard deviation and the lowest R2 value. Intrachip replicates demonstrate the greatest R^2 value and the smallest standard deviation.

Mean group R^2 values were obtained by averaging the correlation value of replicate pairs per group. An additional group representing random non-replicate pairings, with one representative per sample group, was added for comparison.

Replicate β Differences between Batch, Chip, and within Chip

We evaluated and quantified technical variation for interbatch, interchip and intrachip replicates. Our analysis is based on the absolute difference in average beta: $|\beta_{replicate1} - \beta_{replicate2}|$ an absolute quantification of signal difference between replicate readings at each probe. We have 3 pairs of replicates from 2 individuals for interbatch comparison, 6 pairs of replicates from 5 individuals for interchip comparison, and 6 pairs of replicates from 6 individuals for for intrachip comparison.

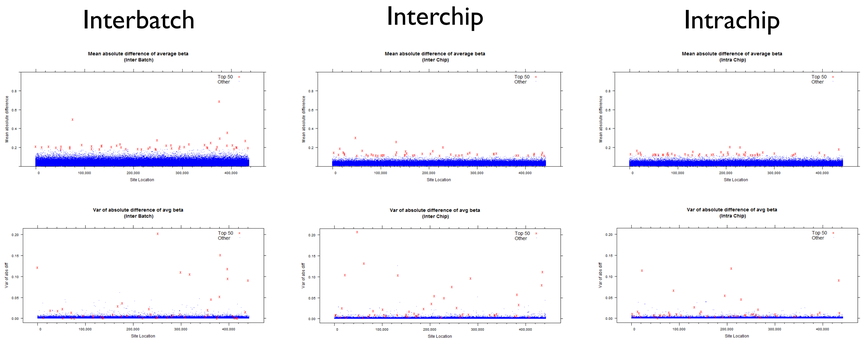


Figure 4. Mean difference in methylation by probe (top row), and variance of differences between replicates (bottom row), plotted by comparison group. For each comparison group, the 50 lowest fidelity probes are identified in red. Differences vary widely across probes. The distribution of the absolute differences has only one mode, despite of the bimodal property of the original average beta (methylated or unmethylated).

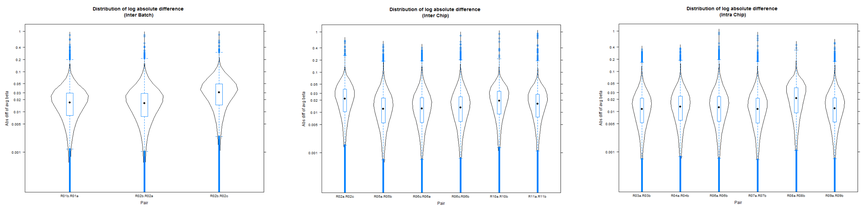


Figure 5. Boxplot and kernel distribution of log absolute methylation difference across 450k sites, for interbatch, interchip and intrachip comparisons. In order to display the body of the distribution as well as outliers, plots are depicted on a log scale for absolute difference.

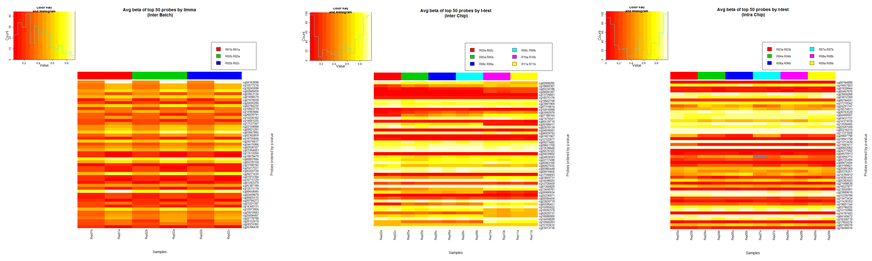


Figure 6. Heatmap of methylation (β) for top 50 variable probes by comparison group. Interchip and intrachip probe differences were compared by one-sided t-tests. Due to sample size limitations, linear modeling was applied for interbatch comparisons. Note that top t-test probes represent small but consistent methylation differences between replicates, rather than the greater overall probe variability.

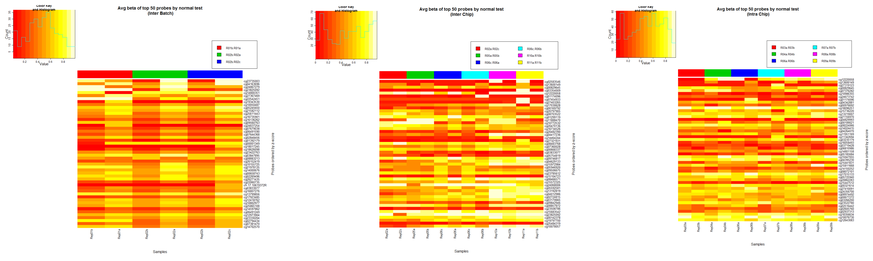


Figure 7. Heatmap of methylation (β) for top 50 variable probes by comparison group, identified by normal test. Note that top normal test probes represent those with greatest overall methylation variability between replicates. Only two probes overlapped between the top 50 probes identified by linear modeling and normal tests of interbatch comparisons. No probes overlapped between the top 50 probes identified by t-tests and normal tests of interchip or intrachip comparisons. Given that probes with the greatest variability generally demonstrate the highest variance (see Figure 4), and thus fail ascertainment by t-test, the absence of overlap is not surprising. Compare methylation values for replicates in top 50 normal test probes to those identified by t-test.

Principal Component Analysis

Principal component analysis was attempted to discover batch effects within the data. The first Principal Component was taken and correlated with the covariates.

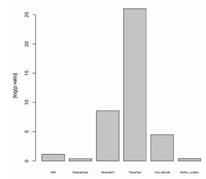


Figure 8. Qualitative visualization of contribution of tissue type to the first principal component. It can be seen that tissue type is a much larger contributor than the random well covariate.

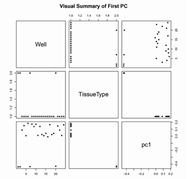
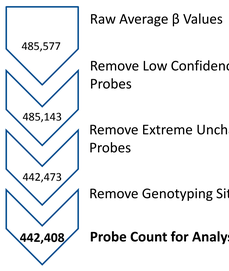


Figure 9. Quantitative visualization of contribution of each covariate to the first principal component. This plot shows that tissue type is a much larger contributor to the first PC, and it is followed by replicate ID and chip barcode.

Quality Control

Prior to analysis of replicate data, raw methylation data (β values) from the HumanMethylation450 array were subjected to typical quality control measures. Although we did not analyze these probes, how many of them were there, and do they have any patterns?



Filtering methodology based on Essex et al. 2011. For the principal component analysis, missing data was imputed using nearest neighbor averaging.

Figure 10. Filtering of raw methylation data. Initial probe count was 485,577 sites per sample. Low confidence probes, or bad data points, were defined as probes/sites with a detection p-value >0.01 and/or a missing average beta value. Rep02a had the highest number of bad probes (3,501) but was not considered an outlier. Bad sites (>25% bad data points across samples) were removed, resulting in a loss of 434 sites, or .0894% of original total. Extreme unchanged sites, with <5% or >95% methylation across all samples, were also removed to increase analytic power, resulting in a loss of 42,670 sites, or 8.787% of the original total. Genotyping sites, meant for unrelated quality control measures, were removed, resulting in a loss of 65 sites, or 0.134% of the original total.

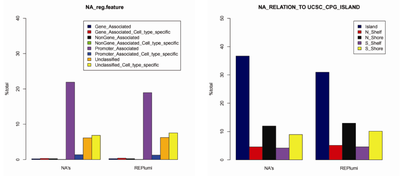


Figure 11. Comparison of filtered, low confidence probe distributions (NA values) to the complete data set, by annotated gene region and CpG island location. Low confidence sites appear to have no preferential distribution versus the full data set.

Conclusions and Prescriptions

How reliable is the Illumina HumanMethylation450 platform? According to their own published technical specifications:

Precision and Accuracy

Reproducibility has been determined based on the correlation of results generated from technical replicates. The HumanMethylation450 BeadChip showed strong correlation between replicates (r>0.98), as well as with the HumanMethylation27 BeadChip and whole-genome bisulfite sequencing (Figure 4).

Here, we have demonstrated that technical replicates actually correlate better than this figure, even across batches a week apart. However, samples run on different batches are somewhat less correlated than samples run in the same batch or on the same chip.

Regarding reliability of methylation values (β values) from probes, Illumina also has published claims:

Sensitivity

By comparing the results of replicate experiments (duplicates of eight biological samples), Illumina scientists have shown that the HumanMethylation450 BeadChip reliably detects a delta-beta value of 0.2 with a lower than 1% false positive rate.

Here again, our results suggest less technical variability than Illumina cautions. From our three interbatch comparisons, only 33 probes demonstrated average absolute $\Delta\beta > 0.2$ out of 485,408 probes. None were significant after multiple test correction. Average differences were less for samples run in the same batch or on the same chip, with no probe differences significant after correction (both Bonferroni and Benjamini-Hochberg methods).