Once the base calls were made for each cycle on both biochemistries, I calculated the error rates for each cycle using the equation $\frac{number\ of\ incorrect\ base\ calls}{total\ number\ of\ nucleotides} \times 100$. I found that the error rates were as follows:

	Biochemistry 1	Biochemistry 2
Cycle 1 Error	3.9%	5.5%
Cycle 2 Error	6.5%	4.9%
Total Error	5.2% ± 1.8	5.2% ± 0.4

While the average error between the two cycles of biochemistries resulted in the same total error of 5.2%, the errors for the two cycles of biochemistry 2 varied significantly less, at a standard deviation of 0.42, or a relative standard deviation of 8.2%. On the other hand, biochemistry 1 resulted in a significantly higher standard deviation between the two cycles, at 1.84, or 35.4% relative standard deviation.

Furthermore, I wanted to investigate how many of these incorrect base calls could be attributed to ambiguous, or "N" reads. I calculated the percent of "N" reads for each of the four cycles and found that all of them resulted in the same amount of ambiguous reads, at 2.0%, leading me to discard this as a source of variation between the two sequencing methods.

Finally, I deemed it to be useful to examine in closer detail, which nucleotides caused the largest amount of error. This was done by first counting the instances of each nucleotide in the reference genome, and then counting how many of the nucleotides were called incorrectly by the software. From this, I was able to calculate the percentage of errors for each of the four nucleotides for both biochemistries 1 and 2. What I found from this analysis was that biochemistry 1 had a significantly high amount of errors reading adenine, at approximately 21% being called incorrectly. On the other hand, biochemistry 2 had a 5.8% error when calling adenine bases. As for the other 3 bases, biochemistry 1 showed better accuracy for base calls compared to biochemistry 2. Additionally, it should also be noted that the average signal intensity for the correct base calls were also found for both biochemistries and included in the summary chart. It appears that the signal intensity for correct base calls in biochemistry 1 was much higher on average, at about 0.5 whereas the average signal intensity for correct base calls in biochemistry 2 was at about 0.3. To explore a possible cause for incorrect base calls, I also included in the right adjacent column, the average signal intensity for the reference base, when they were called incorrectly by the sequencer. I found that biochemistry 1 also had a lower signal intensity for incorrect base calls, at approximately 0.06, whereas the signal intensity for biochemistry 2's incorrect base calls were around 0.12. My findings can be seen in the summary charts below:

Biochem 1

nucleotide	Avg correct signal intensity	Avg incorrect signal intensity	Error counts	Total counts
A	0.499485	0.066370	46	509
C	0.492514	0.082500	4	494

G	0.497853	0.052400	5	509
T	0.497550	0.054667	9	488

Biochem 2

nucleotide	Avg correct signal intensity	Avg incorrect signal intensity	Error counts	Total counts
A	0.297766	0.104200	15	502
C	0.301624	0.119379	29	475
G	0.299207	0.110667	9	512
T	0.295464	0.128182	11	511

Conclusion

Despite both sequencing methods having a similar error rate, I propose the biochemistry team to focus further developments on biochemistry 1. I would suggest that the color reader for adenine on biochemistry 1 undergo troubleshooting to determine the cause for the large source of error, as solving this issue would result in a massive increase in the accuracy of biochemistry 1. Additionally the difference between the average signal intensity of correct and incorrect base calls for biochemistry 1, at 0.44, is much higher than the 0.18 difference for biochemistry 2. This leads me to believe that biochemistry 1's method to be the more accurate of the two due to the larger margin of error for a signal intensity to still be read correctly.