Highly Accurate Chip-based Resequecing of SARS-CoV-2 Clinical Samples
Kendall Hoff1+, Xun Ding1+, Lucas Carter1, John Duque1, Ju-Yu Lin1, Samantha Dung1, Priyanka Singh1, Jiayi Sun1, Filip Crnogorac1, Radha Swaminathan2, Emily N Alden2, Xuechen Zhu2, Ryota Shimada2, Marijan Posavi2, Noah Hull3, Darrell Dinwiddie4, Adam M. Halasz5, Glenn McGall1, Wei Zhou1*, Jeremy S. Edwards2*

1Centrillion Technologies, Palo Alto, CA 94303
2Department of Chemistry and Chemical Biology, University of New Mexico, Albuquerque, NM 87131
3Wyoming Public Health Laboratory, Wyoming Department of Health, Cheyenne, WY 82007
4Department of Pediatrics, University of New Mexico Health Sciences Center, Albuquerque, NM 87131
5Department of Mathematics, West Virginia University, Morgantown, WV, 26506

+Contributed equally
*Correspondence to Jeremy S. Edwards (jsedward@unm.edu) and Wei Zhou (wzhou@centrilliontech.com)

Contents

Number of pages: 66
Number of tables: 1
Number of figures: 32

Figure S1-S24: Density plots for each sample at specific exposure times
Figure S25-S32: Illustrations of sequencing accuracy for each sample
| Recommended Product Name | Vendor | Catalog Number | Lot Number                  |
|--------------------------|--------|----------------|-----------------------------|
| ARTIC nCOV-2019          | IDT    | 100006786 (Pool 1); 100006787 (Pool 2) | 0000514188 (Pool 1); '0000514189 (Pool 2) |
| ARTIC nCOV-2019          | IDT    | 100006786 (Pool 1); 100006787 (Pool 2) | 0000523305 (Pool 1); '000052307 (Pool 2) |
| 2019-nCoV_N1-F           | IDT    | 10006821, 10006830 | 535913                     |
| 2019-nCoV_N1-R           | IDT    | 10006822, 10006831 | 535676                     |
| 2019-nCoV_N2-F           | IDT    | 10006824, 10006833 | 535677                     |
| 2019-nCoV_N2-R           | IDT    | 10006825, 10006834 | 533626                     |
| 2019-nCoV_Rp-F           | IDT    | 10006827, 10006836 | 535679                     |
| 2019-nCoV_Rp-R           | IDT    | 10006828, 10006837 | 535680                     |

**Table S1.** Catalog and lot reference numbers for all primers ordered from Integrated DNA technologies.
**Figure S1. Density plots for WY24 1s scan.** Development of the maximum likelihood base caller for SARS-CoV-2 genome sequencing using full genome tiling arrays. A. 2D histogram of the incorrect calls from all tiling array feature sets including sense and antisense probes for one exposure. This image was constructed by ‘calling’ each base in the genome from all genome features. If the base call does not match the reference, the difference and differential is used to construct the histogram. B. 2D histogram of the correct calls from all tiling array feature sets including sense and antisense probes. This image was constructed by ‘calling’ each base in the genome from all genome features. If the base call matches the reference, the difference and differential were used to construct the 2D histogram. It can be observed that distributions of the difference and differential of ‘correct’ calls is very different from the ‘incorrect’ calls. C. Using the observation from panels A and B, we constructed a function to assign the likelihood that a probe set is calling the correct base for a given position. The dotted contours define the likelihood that the probe set is calling the correct base based on the difference and differential score for that probe set. The triangle points on the plot illustrate the different and differential values for probe sets for all variant sites that have been reported in Wyoming samples in the GISAID database as of August 2020. The green triangles indicate that the base call from this scan suggests a reference call, whereas a red triangle indicates that the call suggests a non-reference base at this position. If the triangle points up, this is from the sense probe, whereas a downward pointed triangle indicates the data is for the antisense probe. The triangle outline is filled in if this probe set from this scan resulted in the highest likelihood for the correct call among all scans.
Figure S2. Density plots for WY24 4s scan. Development of the maximum likelihood base caller for SARS-CoV-2 genome sequencing using full genome tiling arrays. A. 2D histogram of the incorrect calls from all tiling array feature sets including sense and antisense probes for one exposure. This image was constructed by ‘calling’ each base in the genome from all genome features. If the base call does not match the reference, the difference and differential is used to construct the histogram. B. 2D histogram of the correct calls from all tiling array feature sets including sense and antisense probes. This image was constructed by ‘calling’ each base in the genome from all genome features. If the base call matches the reference, the difference and differential were used to construct the 2D histogram. It can be observed that distributions of the difference and differential of ‘correct’ calls is very different from the ‘incorrect’ calls. C. Using the observation from panels A and B, we constructed a function to assign the likelihood that a probe set is calling the correct base for a given position. The dotted contours define the likelihood that the probe set is calling the correct base based on the difference and differential score for that probe set. The triangle points on the plot illustrate the different and differential values for probe sets for all variant sites that have been reported in Wyoming samples in the GISAID database as of August 2020. The green triangles indicate that the base call from this scan suggests a reference call, whereas a red triangle indicates that the call suggests a non-reference base at this position. If the triangle points up, this is from the sense probe, whereas a downward pointed triangle indicates the data is for the antisense probe. The triangle outline is filled in if this probe set from this scan resulted in the highest likelihood for the correct call among all scans.
Figure S3. Density plots for WY24 0.5s scan. Development of the maximum likelihood base caller for SARS-CoV-2 genome sequencing using full genome tiling arrays. A. 2D histogram of the incorrect calls from all tiling array feature sets including sense and antisense probes for one exposure. This image was constructed by ‘calling’ each base in the genome from all genome features. If the base call does not match the reference, the difference and differential is used to construct the histogram. B. 2D histogram of the correct calls from all tiling array feature sets including sense and antisense probes. This image was constructed by ‘calling’ each base in the genome from all genome features. If the base call matches the reference, the difference and differential were used to construct the 2D histogram. It can be observed that distributions of the difference and differential of ‘correct’ calls is very different from the ‘incorrect’ calls. C. Using the observation from panels A and B, we constructed a function to assign the likelihood that a probe set is calling the correct base for a given position. The dotted contours define the likelihood that the probe set is calling the correct base based on the difference and differential score for that probe set. The triangle points on the plot illustrate the different and differential values for probe sets for all variant sites that have been reported in Wyoming samples in the GISAID database as of August 2020. The green triangles indicate that the base call from this scan suggests a reference call, whereas a red triangle indicates that the call suggests a non-reference base at this position. If the triangle points up, this is from the sense probe, whereas a downward pointed triangle indicates the data is for the antisense probe. The triangle outline is filled in if this probe set from this scan resulted in the highest likelihood for the correct call among all scans.
Figure S4. Density plots for WY26 1s scan. Development of the maximum likelihood base caller for SARS-CoV-2 genome sequencing using full genome tiling arrays. A. 2D histogram of the incorrect calls from all tiling array feature sets including sense and antisense probes for one exposure. This image was constructed by ‘calling’ each base in the genome from all genome features. If the base call does not match the reference, the difference and differential is used to construct the histogram. B. 2D histogram of the correct calls from all tiling array feature sets including sense and antisense probes. This image was constructed by ‘calling’ each base in the genome from all genome features. If the base call matches the reference, the difference and differential were used to construct the 2D histogram. It can be observed that distributions of the difference and differential of ‘correct’ calls is very different from the ‘incorrect’ calls. C. Using the observation from panels A and B, we constructed a function to assign the likelihood that a probe set is calling the correct base for a given position. The dotted contours define the likelihood that the probe set is calling the correct base based on the difference and differential score for that probe set. The triangle points on the plot illustrate the different and differential values for probe sets for all variant sites that have been reported in Wyoming samples in the GISAID database as of August 2020. The green triangles indicate that the base call from this scan suggests a reference call, whereas a red triangle indicates that the call suggests a non-reference base at this position. If the triangle points up, this is from the sense probe, whereas a downward pointed triangle indicates the data is for the antisense probe. The triangle outline is filled in if this probe set from this scan resulted in the highest likelihood for the correct call among all scans.
Figure S5. Density plots for WY26 4s scan. Development of the maximum likelihood base caller for SARS-CoV-2 genome sequencing using full genome tiling arrays. A. 2D histogram of the incorrect calls from all tiling array feature sets including sense and antisense probes for one exposure. This image was constructed by ‘calling’ each base in the genome from all genome features. If the base call does not match the reference, the difference and differential is used to construct the histogram. B. 2D histogram of the correct calls from all tiling array feature sets including sense and antisense probes. This image was constructed by ‘calling’ each base in the genome from all genome features. If the base call matches the reference, the difference and differential were used to construct the 2D histogram. It can be observed that distributions of the difference and differential of ‘correct’ calls is very different from the ‘incorrect’ calls. C. Using the observation from panels A and B, we constructed a function to assign the likelihood that a probe set is calling the correct base for a given position. The dotted contours define the likelihood that the probe set is calling the correct base based on the difference and differential score for that probe set. The triangle points on the plot illustrate the different and differential values for probe sets for all variant sites that have been reported in Wyoming samples in the GISAID database as of August 2020. The green triangles indicate that the base call from this scan suggests a reference call, whereas a red triangle indicates that the call suggests a non-reference base at this position. If the triangle points up, this is from the sense probe, whereas a downward pointed triangle indicates the data is for the antisense probe. The triangle outline is filled in if this probe set from this scan resulted in the highest likelihood for the correct call among all scans.
**Figure S6. Density plots for WY26 0.5s scan.** Development of the maximum likelihood base caller for SARS-CoV-2 genome sequencing using full genome tiling arrays. A. 2D histogram of the incorrect calls from all tiling array feature sets including sense and antisense probes for one exposure. This image was constructed by ‘calling’ each base in the genome from all genome features. If the base call does not match the reference, the difference and differential is used to construct the histogram. B. 2D histogram of the correct calls from all tiling array feature sets including sense and antisense probes. This image was constructed by ‘calling’ each base in the genome from all genome features. If the base call matches the reference, the difference and differential were used to construct the 2D histogram. It can be observed that distributions of the difference and differential of ‘correct’ calls is very different from the ‘incorrect’ calls. C. Using the observation from panels A and B, we constructed a function to assign the likelihood that a probe set is calling the correct base for a given position. The dotted contours define the likelihood that the probe set is calling the correct base based on the difference and differential score for that probe set. The triangle points on the plot illustrate the different and differential values for probe sets for all variant sites that have been reported in Wyoming samples in the GISAID database as of August 2020. The green triangles indicate that the base call from this scan suggests a reference call, whereas a red triangle indicates that the call suggests a non-reference base at this position. If the triangle points up, this is from the sense probe, whereas a downward pointed triangle indicates the data is for the antisense probe. The triangle outline is filled in if this probe set from this scan resulted in the highest likelihood for the correct call among all scans.
Figure S7. Density plots for WY32 1s scan. Development of the maximum likelihood base caller for SARS-CoV-2 genome sequencing using full genome tiling arrays. A. 2D histogram of the incorrect calls from all tiling array feature sets including sense and antisense probes for one exposure. This image was constructed by ‘calling’ each base in the genome from all genome features. If the base call does not match the reference, the difference and differential is used to construct the histogram. B. 2D histogram of the correct calls from all tiling array feature sets including sense and antisense probes. This image was constructed by ‘calling’ each base in the genome from all genome features. If the base call matches the reference, the difference and differential were used to construct the 2D histogram. It can be observed that distributions of the difference and differential of ‘correct’ calls is very different from the ‘incorrect’ calls. C. Using the observation from panels A and B, we constructed a function to assign the likelihood that a probe set is calling the correct base for a given position. The dotted contours define the likelihood that the probe set is calling the correct base based on the difference and differential score for that probe set. The triangle points on the plot illustrate the different and differential values for probe sets for all variant sites that have been reported in Wyoming samples in the GISAID database as of August 2020. The green triangles indicate that the base call from this scan suggests a reference call, whereas a red triangle indicates that the call suggests a non-reference base at this position. If the triangle points up, this is from the sense probe, whereas a downward pointed triangle indicates the data is for the antisense probe. The triangle outline is filled in if this probe set from this scan resulted in the highest likelihood for the correct call among all scans.
Figure S8. Density plots for WY32 4s scan. Development of the maximum likelihood base caller for SARS-CoV-2 genome sequencing using full genome tiling arrays. A. 2D histogram of the incorrect calls from all tiling array feature sets including sense and antisense probes for one exposure. This image was constructed by ‘calling’ each base in the genome from all genome features. If the base call does not match the reference, the difference and differential is used to construct the histogram. B. 2D histogram of the correct calls from all tiling array feature sets including sense and antisense probes. This image was constructed by ‘calling’ each base in the genome from all genome features. If the base call matches the reference, the difference and differential were used to construct the 2D histogram. It can be observed that distributions of the difference and differential of ‘correct’ calls is very different from the ‘incorrect’ calls. C. Using the observation from panels A and B, we constructed a function to assign the likelihood that a probe set is calling the correct base for a given position. The dotted contours define the likelihood that the probe set is calling the correct base based on the difference and differential score for that probe set. The triangle points on the plot illustrate the different and differential values for probe sets for all variant sites that have been reported in Wyoming samples in the GISAID database as of August 2020. The green triangles indicate that the base call from this scan suggests a reference call, whereas a red triangle indicates that the call suggests a non-reference base at this position. If the triangle points up, this is from the sense probe, whereas a downward pointed triangle indicates the data is for the antisense probe. The triangle outline is filled in if this probe set from this scan resulted in the highest likelihood for the correct call among all scans.
**Figure S9. Density plots for WY32 0.5s scan.** Development of the maximum likelihood base caller for SARS-CoV-2 genome sequencing using full genome tiling arrays. A. 2D histogram of the incorrect calls from all tiling array feature sets including sense and antisense probes for one exposure. This image was constructed by ‘calling’ each base in the genome from all genome features. If the base call does not match the reference, the difference and differential is used to construct the histogram. B. 2D histogram of the correct calls from all tiling array feature sets including sense and antisense probes. This image was constructed by ‘calling’ each base in the genome from all genome features. If the base call matches the reference, the difference and differential were used to construct the 2D histogram. It can be observed that distributions of the difference and differential of ‘correct’ calls is very different from the ‘incorrect’ calls. C. Using the observation from panels A and B, we constructed a function to assign the likelihood that a probe set is calling the correct base for a given position. The dotted contours define the likelihood that the probe set is calling the correct base based on the difference and differential score for that probe set. The triangle points on the plot illustrate the different and differential values for probe sets for all variant sites that have been reported in Wyoming samples in the GISAID database as of August 2020. The green triangles indicate that the base call from this scan suggests a reference call, whereas a red triangle indicates that the call suggests a non-reference base at this position. If the triangle points up, this is from the sense probe, whereas a downward pointed triangle indicates the data is for the antisense probe. The triangle outline is filled in if this probe set from this scan resulted in the highest likelihood for the correct call among all scans.
Figure S10. Density plots for WY36 1s scan. Development of the maximum likelihood base caller for SARS-CoV-2 genome sequencing using full genome tiling arrays. A. 2D histogram of the incorrect calls from all tiling array feature sets including sense and antisense probes for one exposure. This image was constructed by ‘calling’ each base in the genome from all genome features. If the base call does not match the reference, the difference and differential is used to construct the histogram. B. 2D histogram of the correct calls from all tiling array feature sets including sense and antisense probes. This image was constructed by ‘calling’ each base in the genome from all genome features. If the base call matches the reference, the difference and differential were used to construct the 2D histogram. It can be observed that distributions of the difference and differential of ‘correct’ calls is very different from the ‘incorrect’ calls. C. Using the observation from panels A and B, we constructed a function to assign the likelihood that a probe set is calling the correct base for a given position. The dotted contours define the likelihood that the probe set is calling the correct base based on the difference and differential score for that probe set. The triangle points on the plot illustrate the different and differential values for probe sets for all variant sites that have been reported in Wyoming samples in the GISAID database as of August 2020. The green triangles indicate that the base call from this scan suggests a reference call, whereas a red triangle indicates that the call suggests a non-reference base at this position. If the triangle points up, this is from the sense probe, whereas a downward pointed triangle indicates the data is for the antisense probe. The triangle outline is filled in if this probe set from this scan resulted in the highest likelihood for the correct call among all scans.
**Figure S11. Density plots for WY36 4s scan.** Development of the maximum likelihood base caller for SARS-CoV-2 genome sequencing using full genome tiling arrays. A. 2D histogram of the incorrect calls from all tiling array feature sets including sense and antisense probes for one exposure. This image was constructed by ‘calling’ each base in the genome from all genome features. If the base call does not match the reference, the difference and differential is used to construct the histogram. B. 2D histogram of the correct calls from all tiling array feature sets including sense and antisense probes. This image was constructed by ‘calling’ each base in the genome from all genome features. If the base call matches the reference, the difference and differential were used to construct the 2D histogram. It can be observed that distributions of the difference and differential of ‘correct’ calls is very different from the ‘incorrect’ calls. C. Using the observation from panels A and B, we constructed a function to assign the likelihood that a probe set is calling the correct base for a given position. The dotted contours define the likelihood that the probe set is calling the correct base based on the difference and differential score for that probe set. The triangle points on the plot illustrate the different and differential values for probe sets for all variant sites that have been reported in Wyoming samples in the GISAID database as of August 2020. The green triangles indicate that the base call from this scan suggests a reference call, whereas a red triangle indicates that the call suggests a non-reference base at this position. If the triangle points up, this is from the sense probe, whereas a downward pointed triangle indicates the data is for the antisense probe. The triangle outline is filled in if this probe set from this scan resulted in the highest likelihood for the correct call among all scans.
Figure S12. Density plots for WY36 0.5s scan. Development of the maximum likelihood base caller for SARS-CoV-2 genome sequencing using full genome tiling arrays. A. 2D histogram of the incorrect calls from all tiling array feature sets including sense and antisense probes for one exposure. This image was constructed by ‘calling’ each base in the genome from all genome features. If the base call does not match the reference, the difference and differential is used to construct the histogram. B. 2D histogram of the correct calls from all tiling array feature sets including sense and antisense probes. This image was constructed by ‘calling’ each base in the genome from all genome features. If the base call matches the reference, the difference and differential were used to construct the 2D histogram. It can be observed that distributions of the difference and differential of ‘correct’ calls is very different from the ‘incorrect’ calls. C. Using the observation from panels A and B, we constructed a function to assign the likelihood that a probe set is calling the correct base for a given position. The dotted contours define the likelihood that the probe set is calling the correct base based on the difference and differential score for that probe set. The triangle points on the plot illustrate the different and differential values for probe sets for all variant sites that have been reported in Wyoming samples in the GISAID database as of August 2020. The green triangles indicate that the base call from this scan suggests a reference call, whereas a red triangle indicates that the call suggests a non-reference base at this position. If the triangle points up, this is from the sense probe, whereas a downward pointed triangle indicates the data is for the antisense probe. The triangle outline is filled in if this probe set from this scan resulted in the highest likelihood for the correct call among all scans.
Figure S13. Density plots for WY41 1s scan. Development of the maximum likelihood base caller for SARS-CoV-2 genome sequencing using full genome tiling arrays. A. 2D histogram of the incorrect calls from all tiling array feature sets including sense and antisense probes for one exposure. This image was constructed by ‘calling’ each base in the genome from all genome features. If the base call does not match the reference, the difference and differential is used to construct the histogram. B. 2D histogram of the correct calls from all tiling array feature sets including sense and antisense probes. This image was constructed by ‘calling’ each base in the genome from all genome features. If the base call matches the reference, the difference and differential were used to construct the 2D histogram. It can be observed that distributions of the difference and differential of ‘correct’ calls is very different from the ‘incorrect’ calls. C. Using the observation from panels A and B, we constructed a function to assign the likelihood that a probe set is calling the correct base for a given position. The dotted contours define the likelihood that the probe set is calling the correct base based on the difference and differential score for that probe set. The triangle points on the plot illustrate the different and differential values for probe sets for all variant sites that have been reported in Wyoming samples in the GISAID database as of August 2020. The green triangles indicate that the base call from this scan suggests a reference call, whereas a red triangle indicates that the call suggests a non-reference base at this position. If the triangle points up, this is from the sense probe, whereas a downward pointed triangle indicates the data is for the antisense probe. The triangle outline is filled in if this probe set from this scan resulted in the highest likelihood for the correct call among all scans.
**Figure S14. Density plots for WY41 4s scan.** Development of the maximum likelihood base caller for SARS-CoV-2 genome sequencing using full genome tiling arrays. A. 2D histogram of the incorrect calls from all tiling array feature sets including sense and antisense probes for one exposure. This image was constructed by ‘calling’ each base in the genome from all genome features. If the base call does not match the reference, the difference and differential is used to construct the histogram. B. 2D histogram of the correct calls from all tiling array feature sets including sense and antisense probes. This image was constructed by ‘calling’ each base in the genome from all genome features. If the base call matches the reference, the difference and differential were used to construct the 2D histogram. It can be observed that distributions of the difference and differential of ‘correct’ calls is very different from the ‘incorrect’ calls. C. Using the observation from panels A and B, we constructed a function to assign the likelihood that a probe set is calling the correct base for a given position. The dotted contours define the likelihood that the probe set is calling the correct base based on the difference and differential score for that probe set. The triangle points on the plot illustrate the different and differential values for probe sets for all variant sites that have been reported in Wyoming samples in the GISAID database as of August 2020. The green triangles indicate that the base call from this scan suggests a reference call, whereas a red triangle indicates that the call suggests a non-reference base at this position. If the triangle points up, this is from the sense probe, whereas a downward pointed triangle indicates the data is for the antisense probe. The triangle outline is filled in if this probe set from this scan resulted in the highest likelihood for the correct call among all scans.
Figure S15. Density plots for WY41 0.5s scan. Development of the maximum likelihood base caller for SARS-CoV-2 genome sequencing using full genome tiling arrays. A. 2D histogram of the incorrect calls from all tiling array feature sets including sense and antisense probes for one exposure. This image was constructed by ‘calling’ each base in the genome from all genome features. If the base call does not match the reference, the difference and differential is used to construct the histogram. B. 2D histogram of the correct calls from all tiling array feature sets including sense and antisense probes. This image was constructed by ‘calling’ each base in the genome from all genome features. If the base call matches the reference, the difference and differential were used to construct the 2D histogram. It can be observed that distributions of the difference and differential of ‘correct’ calls is very different from the ‘incorrect’ calls. C. Using the observation from panels A and B, we constructed a function to assign the likelihood that a probe set is calling the correct base for a given position. The dotted contours define the likelihood that the probe set is calling the correct base based on the difference and differential score for that probe set. The triangle points on the plot illustrate the different and differential values for probe sets for all variant sites that have been reported in Wyoming samples in the GISAID database as of August 2020. The green triangles indicate that the base call from this scan suggests a reference call, whereas a red triangle indicates that the call suggests a non-reference base at this position. If the triangle points up, this is from the sense probe, whereas a downward pointed triangle indicates the data is for the antisense probe. The triangle outline is filled in if this probe set from this scan resulted in the highest likelihood for the correct call among all scans.
Figure S16. Density plots for WY44 1s scan. Development of the maximum likelihood base caller for SARS-CoV-2 genome sequencing using full genome tiling arrays. A. 2D histogram of the incorrect calls from all tiling array feature sets including sense and antisense probes for one exposure. This image was constructed by ‘calling’ each base in the genome from all genome features. If the base call does not match the reference, the difference and differential is used to construct the histogram. B. 2D histogram of the correct calls from all tiling array feature sets including sense and antisense probes. This image was constructed by ‘calling’ each base in the genome from all genome features. If the base call matches the reference, the difference and differential were used to construct the 2D histogram. It can be observed that distributions of the difference and differential of ‘correct’ calls is very different from the ‘incorrect’ calls. C. Using the observation from panels A and B, we constructed a function to assign the likelihood that a probe set is calling the correct base for a given position. The dotted contours define the likelihood that the probe set is calling the correct base based on the difference and differential score for that probe set. The triangle points on the plot illustrate the different and differential values for probe sets for all variant sites that have been reported in Wyoming samples in the GISAID database as of August 2020. The green triangles indicate that the base call from this scan suggests a reference call, whereas a red triangle indicates that the call suggests a non-reference base at this position. If the triangle points up, this is from the sense probe, whereas a downward pointed triangle indicates the data is for the antisense probe. The triangle outline is filled in if this probe set from this scan resulted in the highest likelihood for the correct call among all scans.
Figure S17. Density plots for WY44 4s scan. Development of the maximum likelihood base caller for SARS-CoV-2 genome sequencing using full genome tiling arrays. A. 2D histogram of the incorrect calls from all tiling array feature sets including sense and antisense probes for one exposure. This image was constructed by ‘calling’ each base in the genome from all genome features. If the base call does not match the reference, the difference and differential is used to construct the histogram. B. 2D histogram of the correct calls from all tiling array feature sets including sense and antisense probes. This image was constructed by ‘calling’ each base in the genome from all genome features. If the base call matches the reference, the difference and differential were used to construct the 2D histogram. It can be observed that distributions of the difference and differential of ‘correct’ calls is very different from the ‘incorrect’ calls. C. Using the observation from panels A and B, we constructed a function to assign the likelihood that a probe set is calling the correct base for a given position. The dotted contours define the likelihood that the probe set is calling the correct base based on the difference and differential score for that probe set. The triangle points on the plot illustrate the different and differential values for probe sets for all variant sites that have been reported in Wyoming samples in the GISAID database as of August 2020. The green triangles indicate that the base call from this scan suggests a reference call, whereas a red triangle indicates that the call suggests a non-reference base at this position. If the triangle points up, this is from the sense probe, whereas a downward pointed triangle indicates the data is for the antisense probe. The triangle outline is filled in if this probe set from this scan resulted in the highest likelihood for the correct call among all scans.
**Figure S18. Density plots for WY44 0.5s scan.** Development of the maximum likelihood base caller for SARS-CoV-2 genome sequencing using full genome tiling arrays. A. 2D histogram of the incorrect calls from all tiling array feature sets including sense and antisense probes for one exposure. This image was constructed by ‘calling’ each base in the genome from all genome features. If the base call does not match the reference, the difference and differential is used to construct the histogram. B. 2D histogram of the correct calls from all tiling array feature sets including sense and antisense probes. This image was constructed by ‘calling’ each base in the genome from all genome features. If the base call matches the reference, the difference and differential were used to construct the 2D histogram. It can be observed that distributions of the difference and differential of ‘correct’ calls is very different from the ‘incorrect’ calls. C. Using the observation from panels A and B, we constructed a function to assign the likelihood that a probe set is calling the correct base for a given position. The dotted contours define the likelihood that the probe set is calling the correct base based on the difference and differential score for that probe set. The triangle points on the plot illustrate the different and differential values for probe sets for all variant sites that have been reported in Wyoming samples in the GISAID database as of August 2020. The green triangles indicate that the base call from this scan suggests a reference call, whereas a red triangle indicates that the call suggests a non-reference base at this position. If the triangle points up, this is from the sense probe, whereas a downward pointed triangle indicates the data is for the antisense probe. The triangle outline is filled in if this probe set from this scan resulted in the highest likelihood for the correct call among all scans.
Figure S19. Density plots for WY59 1s scan. Development of the maximum likelihood base caller for SARS-CoV-2 genome sequencing using full genome tiling arrays. A. 2D histogram of the incorrect calls from all tiling array feature sets including sense and antisense probes for one exposure. This image was constructed by ‘calling’ each base in the genome from all genome features. If the base call does not match the reference, the difference and differential is used to construct the histogram. B. 2D histogram of the correct calls from all tiling array feature sets including sense and antisense probes. This image was constructed by ‘calling’ each base in the genome from all genome features. If the base call matches the reference, the difference and differential were used to construct the 2D histogram. It can be observed that distributions of the difference and differential of ‘correct’ calls is very different from the ‘incorrect’ calls. C. Using the observation from panels A and B, we constructed a function to assign the likelihood that a probe set is calling the correct base for a given position. The dotted contours define the likelihood that the probe set is calling the correct base based on the difference and differential score for that probe set. The triangle points on the plot illustrate the different and differential values for probe sets for all variant sites that have been reported in Wyoming samples in the GISAID database as of August 2020. The green triangles indicate that the base call from this scan suggests a reference call, whereas a red triangle indicates that the call suggests a non-reference base at this position. If the triangle points up, this is from the sense probe, whereas a downward pointed triangle indicates the data is for the antisense probe. The triangle outline is filled in if this probe set from this scan resulted in the highest likelihood for the correct call among all scans.
**Figure S20. Density plots for WY59 4s scan.** Development of the maximum likelihood base caller for SARS-CoV-2 genome sequencing using full genome tiling arrays. A. 2D histogram of the incorrect calls from all tiling array feature sets including sense and antisense probes for one exposure. This image was constructed by ‘calling’ each base in the genome from all genome features. If the base call does not match the reference, the difference and differential is used to construct the histogram. B. 2D histogram of the correct calls from all tiling array feature sets including sense and antisense probes. This image was constructed by ‘calling’ each base in the genome from all genome features. If the base call matches the reference, the difference and differential were used to construct the 2D histogram. It can be observed that distributions of the difference and differential of ‘correct’ calls is very different from the ‘incorrect’ calls. C. Using the observation from panels A and B, we constructed a function to assign the likelihood that a probe set is calling the correct base for a given position. The dotted contours define the likelihood that the probe set is calling the correct base based on the difference and differential score for that probe set. The triangle points on the plot illustrate the different and differential values for probe sets for all variant sites that have been reported in Wyoming samples in the GISAID database as of August 2020. The green triangles indicate that the base call from this scan suggests a reference call, whereas a red triangle indicates that the call suggests a non-reference base at this position. If the triangle points up, this is from the sense probe, whereas a downward pointed triangle indicates the data is for the antisense probe. The triangle outline is filled in if this probe set from this scan resulted in the highest likelihood for the correct call among all scans.
Figure S21. Density plots for WY59 0.5s scan. Development of the maximum likelihood base caller for SARS-CoV-2 genome sequencing using full genome tiling arrays. A. 2D histogram of the incorrect calls from all tiling array feature sets including sense and antisense probes for one exposure. This image was constructed by ‘calling’ each base in the genome from all genome features. If the base call does not match the reference, the difference and differential is used to construct the histogram. B. 2D histogram of the correct calls from all tiling array feature sets including sense and antisense probes. This image was constructed by ‘calling’ each base in the genome from all genome features. If the base call matches the reference, the difference and differential were used to construct the 2D histogram. It can be observed that distributions of the difference and differential of ‘correct’ calls is very different from the ‘incorrect’ calls. C. Using the observation from panels A and B, we constructed a function to assign the likelihood that a probe set is calling the correct base for a given position. The dotted contours define the likelihood that the probe set is calling the correct base based on the difference and differential score for that probe set. The triangle points on the plot illustrate the different and differential values for probe sets for all variant sites that have been reported in Wyoming samples in the GISAID database as of August 2020. The green triangles indicate that the base call from this scan suggests a reference call, whereas a red triangle indicates that the call suggests a non-reference base at this position. If the triangle points up, this is from the sense probe, whereas a downward pointed triangle indicates the data is for the antisense probe. The triangle outline is filled in if this probe set from this scan resulted in the highest likelihood for the correct call among all scans.
**Figure S22. Density plots for WY64 1s scan.** Development of the maximum likelihood base caller for SARS-CoV-2 genome sequencing using full genome tiling arrays. A. 2D histogram of the incorrect calls from all tiling array feature sets including sense and antisense probes for one exposure. This image was constructed by ‘calling’ each base in the genome from all genome features. If the base call does not match the reference, the difference and differential is used to construct the histogram. B. 2D histogram of the correct calls from all tiling array feature sets including sense and antisense probes. This image was constructed by ‘calling’ each base in the genome from all genome features. If the base call matches the reference, the difference and differential were used to construct the 2D histogram. It can be observed that distributions of the difference and differential of ‘correct’ calls is very different from the ‘incorrect’ calls. C. Using the observation from panels A and B, we constructed a function to assign the likelihood that a probe set is calling the correct base for a given position. The dotted contours define the likelihood that the probe set is calling the correct base based on the difference and differential score for that probe set. The triangle points on the plot illustrate the different and differential values for probe sets for all variant sites that have been reported in Wyoming samples in the GISAID database as of August 2020. The green triangles indicate that the base call from this scan suggests a reference call, whereas a red triangle indicates that the call suggests a non-reference base at this position. If the triangle points up, this is from the sense probe, whereas a downward pointed triangle indicates the data is for the antisense probe. The triangle outline is filled in if this probe set from this scan resulted in the highest likelihood for the correct call among all scans.
**Figure S23. Density plots for WY64 4s scan.** Development of the maximum likelihood base caller for SARS-CoV-2 genome sequencing using full genome tiling arrays. A. 2D histogram of the incorrect calls from all tiling array feature sets including sense and antisense probes for one exposure. This image was constructed by ‘calling’ each base in the genome from all genome features. If the base call does not match the reference, the difference and differential is used to construct the histogram. B. 2D histogram of the correct calls from all tiling array feature sets including sense and antisense probes. This image was constructed by ‘calling’ each base in the genome from all genome features. If the base call matches the reference, the difference and differential were used to construct the 2D histogram. It can be observed that distributions of the difference and differential of ‘correct’ calls is very different from the ‘incorrect’ calls. C. Using the observation from panels A and B, we constructed a function to assign the likelihood that a probe set is calling the correct base for a given position. The dotted contours define the likelihood that the probe set is calling the correct base based on the difference and differential score for that probe set. The triangle points on the plot illustrate the different and differential values for probe sets for all variant sites that have been reported in Wyoming samples in the GISAID database as of August 2020. The green triangles indicate that the base call from this scan suggests a reference call, whereas a red triangle indicates that the call suggests a non-reference base at this position. If the triangle points up, this is from the sense probe, whereas a downward pointed triangle indicates the data is for the antisense probe. The triangle outline is filled in if this probe set from this scan resulted in the highest likelihood for the correct call among all scans.
**Figure S24. Density plots for WY64 0.5s scan.** Development of the maximum likelihood base caller for SARS-CoV-2 genome sequencing using full genome tiling arrays. A. 2D histogram of the incorrect calls from all tiling array feature sets including sense and antisense probes for one exposure. This image was constructed by ‘calling’ each base in the genome from all genome features. If the base call does not match the reference, the difference and differential is used to construct the histogram. B. 2D histogram of the correct calls from all tiling array feature sets including sense and antisense probes. This image was constructed by ‘calling’ each base in the genome from all genome features. If the base call matches the reference, the difference and differential were used to construct the 2D histogram. It can be observed that distributions of the difference and differential of ‘correct’ calls is very different from the ‘incorrect’ calls. C. Using the observation from panels A and B, we constructed a function to assign the likelihood that a probe set is calling the correct base for a given position. The dotted contours define the likelihood that the probe set is calling the correct base based on the difference and differential score for that probe set. The triangle points on the plot illustrate the different and differential values for probe sets for all variant sites that have been reported in Wyoming samples in the GISAID database as of August 2020. The green triangles indicate that the base call from this scan suggests a reference call, whereas a red triangle indicates that the call suggests a non-reference base at this position. If the triangle points up, this is from the sense probe, whereas a downward pointed triangle indicates the data is for the antisense probe. The triangle outline is filled in if this probe set from this scan resulted in the highest likelihood for the correct call among all scans.
Figure S25. Sequencing accuracy across the SARS-CoV-2 genome for WY24. The Phred score (left axis) for all bases in the SARS-CoV-2 genome from the tiling array full genome sequencing of WY24. The positions of all variant calls are highlighted by Black “X,” and a Red “X” indicates this is a correct variant call (confirmed by the Illumina short read sequencing data). The cumulative sum of non-calls (Blue line), variant calls with a Phred score greater than 20 (Cyan line), and variant calls that have a Phred score greater than 20 and pass the low coverage filter (Red line). B. Comparison of the tiling array genome sequencing quality scores and variant calls to the amplicon coverage from short read Illumina sequencing data. The light blue (right axis) lines indicate the sequence coverage from the WY24 sample, and the dark blue lines indicate the average sequencing coverage over all Wyoming GISAID samples as of 8/2020.
Figure S26. Sequencing accuracy across the SARS-CoV-2 genome for WY26.
The Phred score (left axis) for all bases in the SARS-CoV-2 genome from the tiling array full
genome sequencing of WY26. The positions of all variant calls are highlighted by Black “X,”
and a Red “X” indicates this is a correct variant call (confirmed by the Illumina short read
sequencing data). The cumulative sum of non-calls (Blue line), variant calls with a Phred score
greater than 20 (Cyan line), and variant calls that have a Phred score greater than 20 and pass the
low coverage filter (Red line). B. Comparison of the tiling array genome sequencing quality
scores and variant calls to the amplicon coverage from short read Illumina sequencing data. The
light blue (right axis) lines indicate the sequence coverage from the WY26 sample, and the dark
blue lines indicate the average sequencing coverage overall Wyoming GISAID samples as of
8/2020.
**Figure S27. Sequencing accuracy across the SARS-CoV-2 genome for WY32.**
The Phred score (left axis) for all bases in the SARS-CoV-2 genome from the tiling array full
genome sequencing of WY32. The positions of all variant calls are highlighted by Black “X,”
and a Red “X” indicates this is a correct variant call (confirmed by the Illumina short read
sequencing data). The cumulative sum of non-calls (Blue line), variant calls with a Phred score
greater than 20 (Cyan line), and variant calls that have a Phred score greater than 20 and pass the
low coverage filter (Red line). B. Comparison of the tiling array genome sequencing quality
scores and variant calls to the amplicon coverage from short read Illumina sequencing data. The
light blue (right axis) lines indicate the sequence coverage from the WY32 sample, and the dark
blue lines indicate the average sequencing coverage over all Wyoming GISAID samples as of
8/2020.
**Figure S28. Sequencing accuracy across the SARS-CoV-2 genome for WY36.**
The Phred score (left axis) for all bases in the SARS-CoV-2 genome from the tiling array full genome sequencing of WY36. The positions of all variant calls are highlighted by Black “X,” and a Red “X” indicates this is a correct variant call (confirmed by the Illumina short read sequencing data). The cumulative sum of non-calls (Blue line), variant calls with a Phred score greater than 20 (Cyan line), and variant calls that have a Phred score greater than 20 and pass the low coverage filter (Red line). B. Comparison of the tiling array genome sequencing quality scores and variant calls to the amplicon coverage from short read Illumina sequencing data. The light blue (right axis) lines indicate the sequence coverage from the WY36 sample, and the dark blue lines indicate the average sequencing coverage over all Wyoming GISAID samples as of 8/2020.
**Figure S29. Sequencing accuracy across the SARS-CoV-2 genome for WY41.**
The Phred score (left axis) for all bases in the SARS-CoV-2 genome from the tiling array full genome sequencing of WY41. The positions of all variant calls are highlighted by Black “X,” and a Red “X” indicates this is a correct variant call (confirmed by the Illumina short read sequencing data). The cumulative sum of non-calls (Blue line), variant calls with a Phred score greater than 20 (Cyan line), and variant calls that have a Phred score greater than 20 and pass the low coverage filter (Red line). B. Comparison of the tiling array genome sequencing quality scores and variant calls to the amplicon coverage from short read Illumina sequencing data. The light blue (right axis) lines indicate the sequence coverage from the WY41 sample, and the dark blue lines indicate the average sequencing coverage over all Wyoming GISAID samples as of 8/2020.
Figure S30. Sequencing accuracy across the SARS-CoV-2 genome for WY44.
The Phred score (left axis) for all bases in the SARS-CoV-2 genome from the tiling array full genome sequencing of WY44. The positions of all variant calls are highlighted by Black “X,” and a Red “X” indicates this is a correct variant call (confirmed by the Illumina short read sequencing data). The cumulative sum of non-calls (Blue line), variant calls with a Phred score greater than 20 (Cyan line), and variant calls that have a Phred score greater than 20 and pass the low coverage filter (Red line). B. Comparison of the tiling array genome sequencing quality scores and variant calls to the amplicon coverage from short read Illumina sequencing data. The light blue (right axis) lines indicate the sequence coverage from the WY44 sample, and the dark blue lines indicate the average sequencing coverage over all Wyoming GISAID samples as of 8/2020.
Figure S31. Sequencing accuracy across the SARS-CoV-2 genome for WY59. The Phred score (left axis) for all bases in the SARS-CoV-2 genome from the tiling array full genome sequencing of WY59. The positions of all variant calls are highlighted by Black “X,” and a Red “X” indicates this is a correct variant call (confirmed by the Illumina short read sequencing data). The cumulative sum of non-calls (Blue line), variant calls with a Phred score greater than 20 (Cyan line), and variant calls that have a Phred score greater than 20 and pass the low coverage filter (Red line). B. Comparison of the tiling array genome sequencing quality scores and variant calls to the amplicon coverage from short read Illumina sequencing data. The light blue (right axis) lines indicate the sequence coverage from the WY59 sample, and the dark blue lines indicate the average sequencing coverage over all Wyoming GISAID samples as of 8/2020.
Figure S32. Sequencing accuracy across the SARS-CoV-2 genome for WY64.
The Phred score (left axis) for all bases in the SARS-CoV-2 genome from the tiling array full genome sequencing of WY62. The positions of all variant calls are highlighted by Black “X,” and a Red “X” indicates this is a correct variant call (confirmed by the Illumina short read sequencing data). The cumulative sum of non-calls (Blue line), variant calls with a Phred score greater than 20 (Cyan line), and variant calls that have a Phred score greater than 20 and pass the low coverage filter (Red line). B. Comparison of the tiling array genome sequencing quality scores and variant calls to the amplicon coverage from short read Illumina sequencing data. The light blue (right axis) lines indicate the sequence coverage from the WY62 sample, and the dark blue lines indicate the average sequencing coverage over all Wyoming GISAID samples as of 8/2020.
