TECHNOLOGY COMPARISON IGV PILEUP

To compare the accuracy and read length (and thus mappability) of different sequencing technologies, PacBio generated an IGV pileup of the human *STRC* locus using alignments of HG002 reads that are publicly available from Genome in a Bottle (https://bit.ly/2ZA4djV).

Datasets

- Illumina 2×250 bp reads from HiSeq 2500 https://bit.ly/3jbIzLN
- PacBio 15+20 kb HiFi reads from Sequel II System, Chemistry 2.0 https://bit.ly/2Wtt8DR
- ONT: Ultralong reads, R9.4.1 guppy V3.2.4 "2020-01-22" https://bit.ly/3eAa2Cl
  - Updated 2020-07-15 to R9.4.1 guppy V3.4.5 "2020-02-04" dataset https://bit.ly/2WrlokI

Visualization in IGV

IGV provides different parameters to suppress/show variation and errors in reads. Interpretation is often impacted by which information is hidden and shown. Since our goal is to show read-level accuracy, the pileup uses:

- Mapping quality ≥20
- Filtered secondary alignments
- Hidden indels <10 bp
- Disabled "Quick consensus mode"
- Coverage allele-fraction threshold: 0.125 with "quality weight allele fraction" enabled
- Enabled "Shade mismatched bases by quality" with max transparency at 0 and none at 20.

IGV PARAMETER SELECTION

To illustrate the impact of IGV parameter selection, the following slides show HiFi reads and ONT reads with different IGV parameter combinations.

Note that the accurate HiFi reads look very similar across views, while hiding/de-emphasizing errors changes the look of the ONT reads.

- "Quick consensus mode" enabled and disabled
- "Quality weight allele fraction" enabled and disabled
- "Shade mismatched bases by quality" enabled and disabled
  - Transparency between base quality 0-20 and 10-30

"Quick consensus mode"
Disabled

"Quality weight allele fraction"
Disabled

"Shade mismatched bases by quality"
Disabled

This view shows all errors in individual reads. We initially used pileups with these settings.
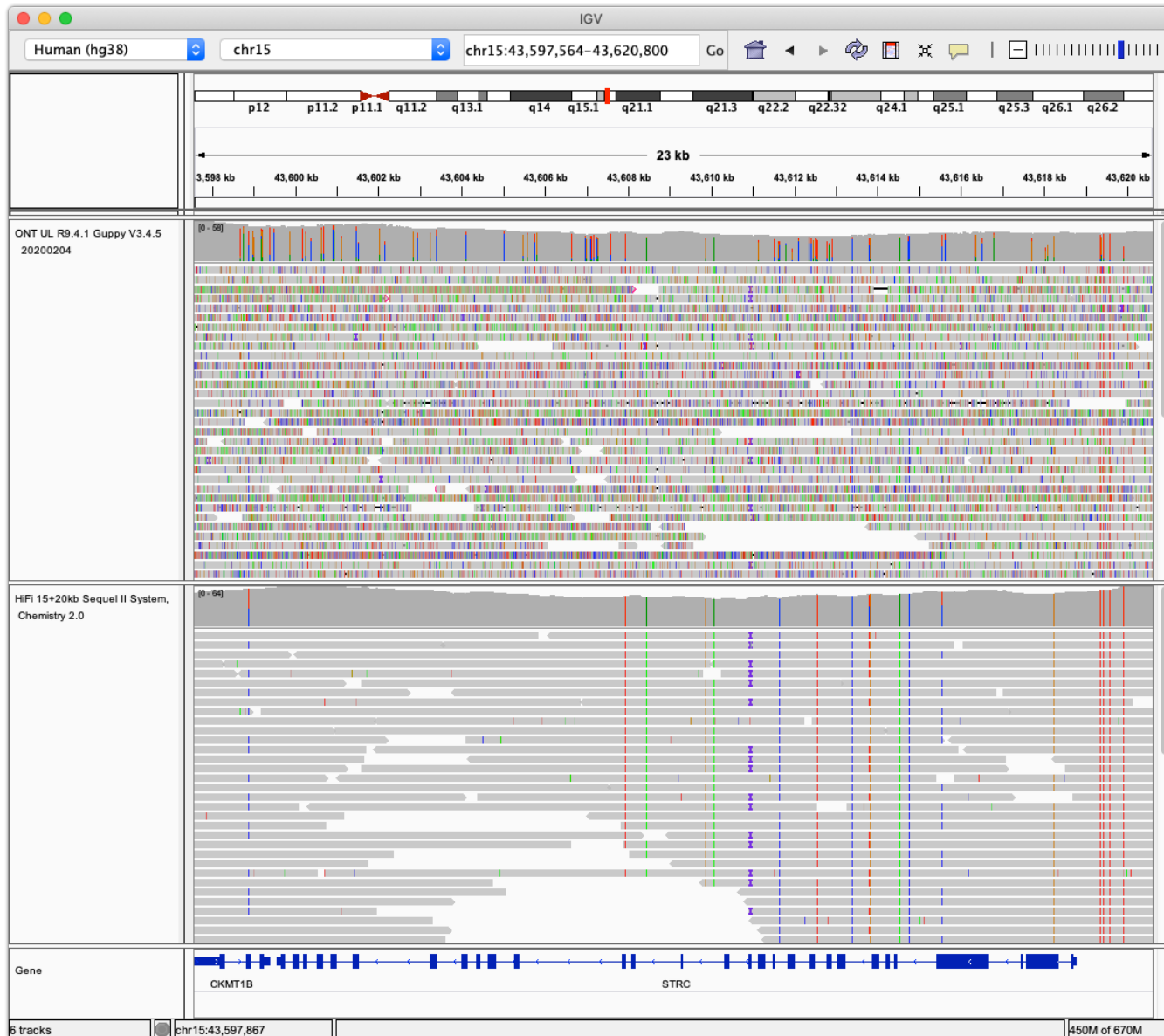

ONT UL R9.4.1

Guppy 3.4.5

20200204

15+20 kb HiFi

Sequel II System

Chemistry 2.0

"Quick consensus mode"
Disabled

"Quality weight allele fraction"
Enabled

"Shade mismatched bases by quality"
Disabled

ONT UL R9.4.1

Guppy 3.4.5

20200204

15+20 kb HiFi

Sequel II System

Chemistry 2.0

"Quick consensus mode"
Enabled

"Quality weight allele fraction"
Enabled

"Shade mismatched bases by quality"
Disabled

This view hides "private" errors seen in individual reads. It is a single "multi-molecule" correction.

ONT UL R9.4.1

Guppy 3.4.5

20200204

15+20 kb HiFi

Sequel II System

Chemistry 2.0

"Quick consensus mode"
Disabled

"Quality weight allele fraction"
Enabled

"Shade mismatched bases by quality"
Enabled. Shade BQ 0-20.

This view de-emphasizes mismatches. Based on community feedback, we will use these settings going forward.

ONT UL R9.4.1
Guppy 3.4.5
20200204

15+20 kb HiFi
Sequel II System
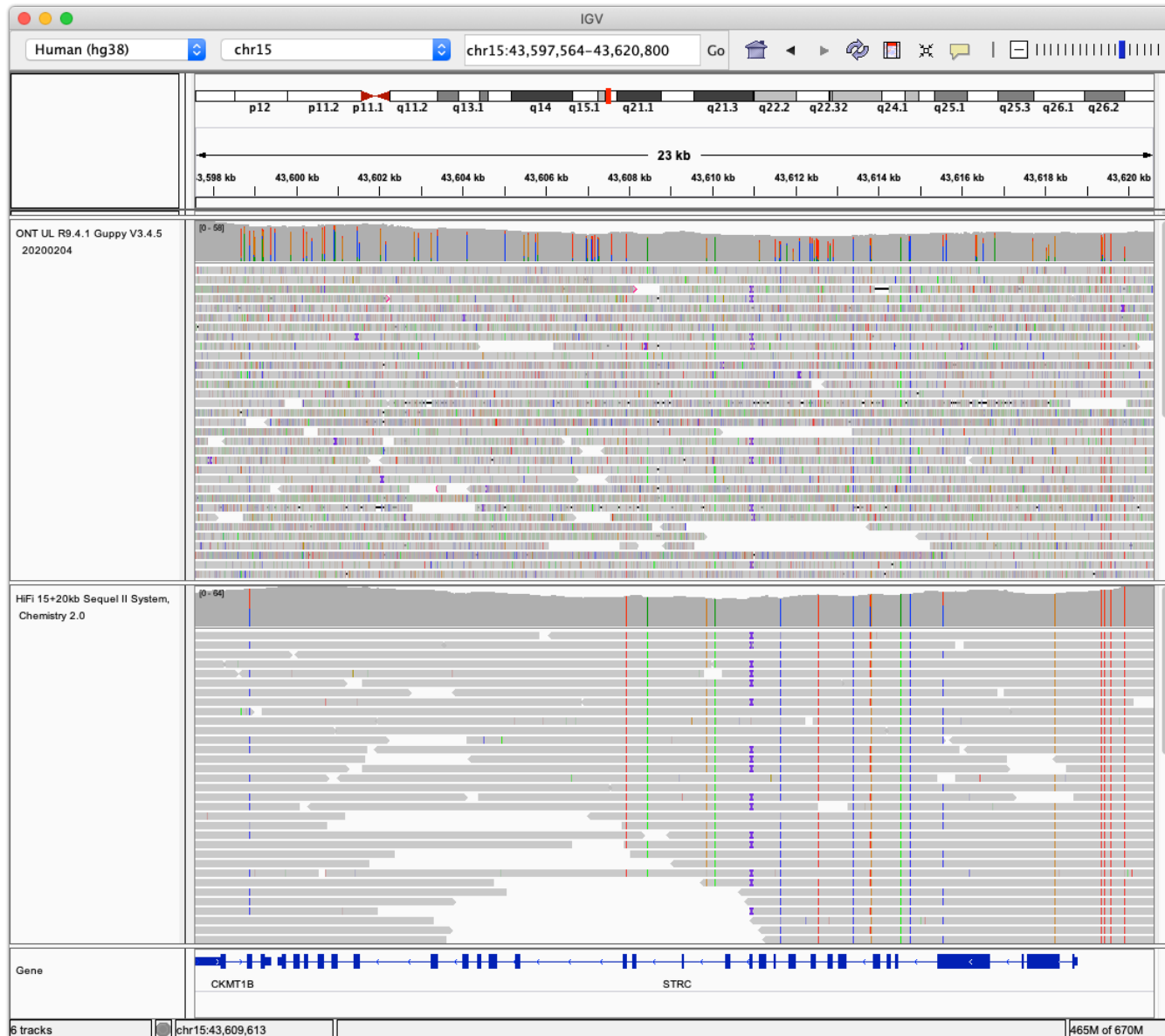Chemistry 2.0

"Quick consensus mode"
Disabled

"Quality weight allele fraction"
Enabled

"Shade mismatched bases by quality"
Enabled. Shade BQ 10-30.

This view aggressively de-emphasizes mismatches.

ONT UL R9.4.1

Guppy 3.4.5

20200204

15+20 kb HiFi

Sequel II System

Chemistry 2.0

www.pacb.com