PacBio

Sequencing by binding (SBB®) delivers unprecedented NGS accuracy

June 8, 2022 | AGBT
Jonas Korlach | CSO

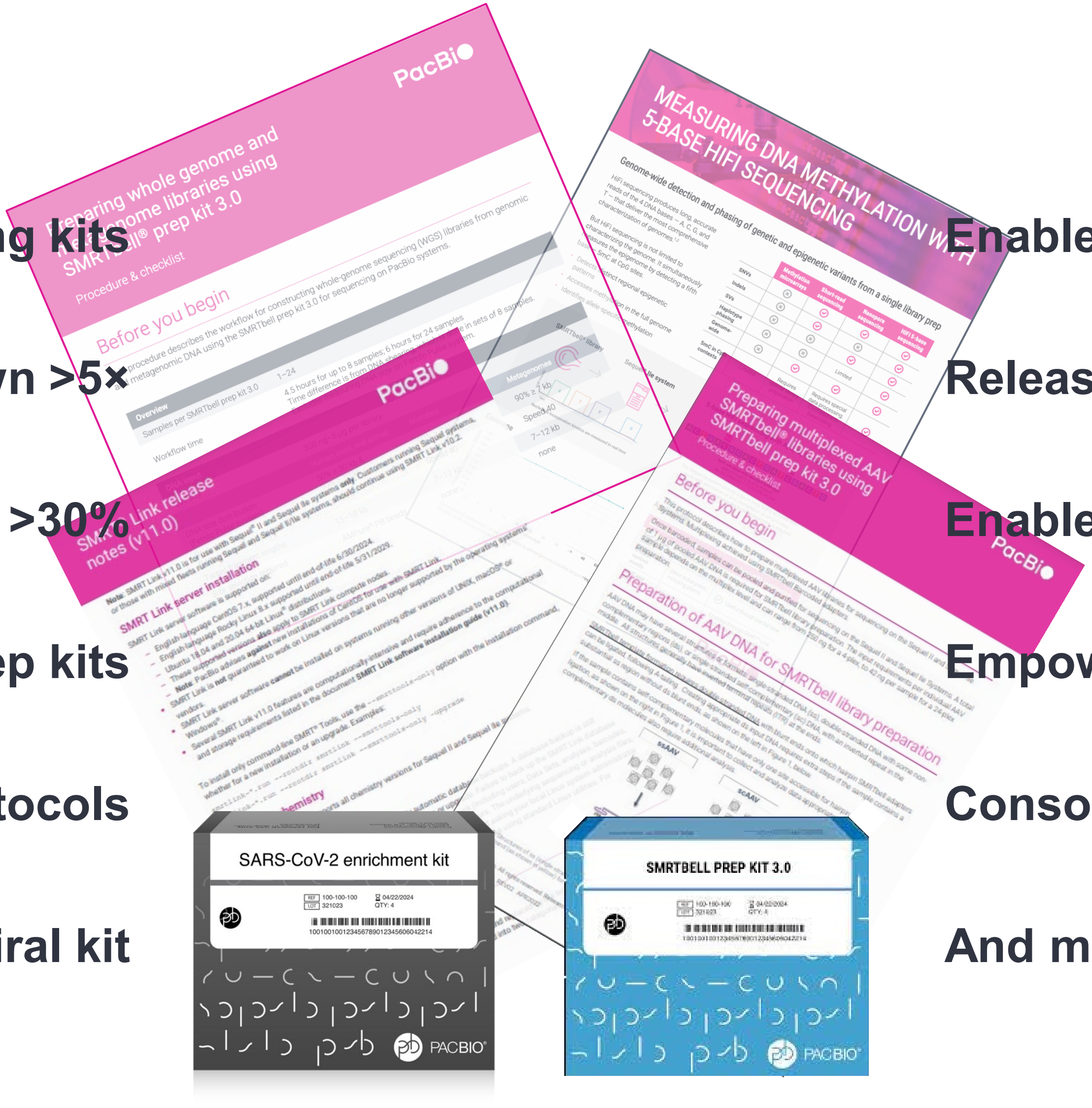
Safe harbor statement

All statements in this presentation (and any accompanying oral presentation) that are not historical are forward-looking statements within the meaning of Section 27A of the Securities Act of 1933, as amended, Section 21E of the Securities Exchange Act of 1934, as amended, and the U.S. Private Securities Litigation Reform Act of 1995, including statements that generally relate to future events or our future financial or operating performance, future applications and availability, release dates, uses, accuracy, advantages, or quality or performance of, or benefits or expected benefits of using, our products or technologies, and the suitability or utility of our products or technologies for particular applications or projects. Such statements are based on management's current beliefs, estimates, assumptions, and projections and on information available to management as of the date of this presentation. Forward-looking statements include, among other things, statements regarding predictions, estimates, plans, and expectations regarding the performance, applications, commercial availability, release dates, uses, accuracy, advantages, quality, performance, and benefits or expected benefits of our products or technologies, and the suitability or utility of our products or technologies for particular applications or projects, including in connection with SBB products, product candidates and technologies. You should not place undue reliance on forward-looking statements because they involve known and unknown risks, uncertainties, changes in circumstances and other factors that are, in some cases, beyond the Company's control and could cause actual results to differ materially from the information expressed or implied by forward- looking statements made in the presentation. Factors that could materially affect actual results can be found in our filings with the Securities and Exchange Commission, including our most recent reports on Forms 8-K, 10-K and 10-Q, and include those listed under the caption "Risk Factors." The Company undertakes no obligation to revise or update information in this presentation to reflect events or circumstances in the future, even if new information becomes available.

The last 18 months have been transformational for PacBio

Genomes	Foundation for T2T, calling all variants
Epigenomes	Simultaneous 5mC calling
Transcriptomes	MAS-Iso-Seq
Chromatin architecture & dynamics	Fiber-Seq/SAMOSA
Metagenomics	Full-length 16S, complete MAGs
Gene therapy	Complete AAV sequencing solution
SARS-CoV-2	HiFiViral

The last 18 months have been transformational for PacBio



- Released new binding kits
- Drove DNA input down >5×
- Increased average yield >30%
- Released new prep kits
- Released automated protocols
- Released HiFiViral kit

- Enabled 5mC calling on instrument
- Released AAV workflow
- Enabled demultiplex on instrument
- Empowered high-throughput processing
- Consolidated workflows and protocols
- And much more…

Accuracy matters — it’s the hallmark of who we are

- **Human genetics — Neuroscience**
- **Human genetics — Immunology**
- **Rare + inherited disease research**
- **Plant + animal sciences**
- **Infectious disease / microbiology**
- **Potential for early-stage cancer screening**
- **Potential for cancer recurrence monitoring**
- **Enabling therapy selection**
- **Targeted clinical panels**
- **Potential for noninvasive prenatal screening**

HiFi sequencing

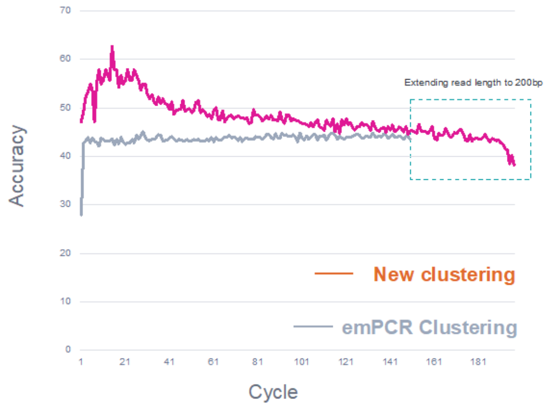
Delivers long reads with the highest accuracy — even in hard-to-sequence regions

SBB sequencing

Promises significant accuracy improvements over conventional NGS approaches

Going beyond what we shared at JP Morgan

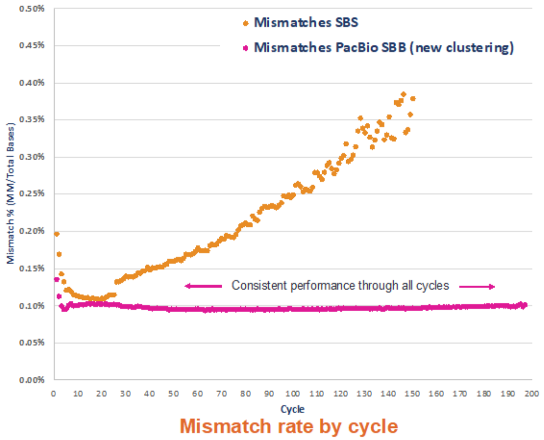
Implemented novel clustering method to vastly improve SBB™ performance



Even lower error rate than original emulsion PCR method

Observed quality scores for internal run - benchmark against lambda

SBB demonstrates greater accuracy & consistency vs current technologies



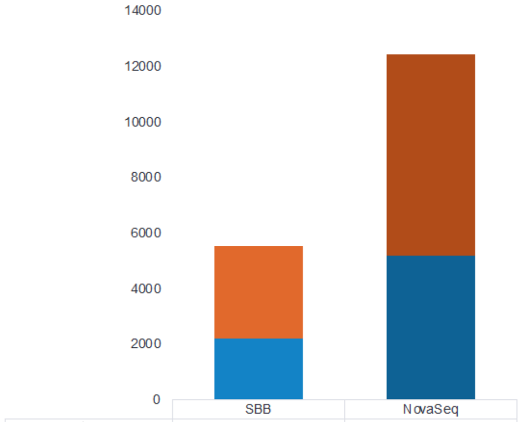
Mismatch rate by cycle

Even lower error rate than original emulsion PCR method

Reaches 0.1% “true variation level” between the sample and reference

Whole genome sequence of HG002 against HG38 reference
SBS sequencing data derived from Precision FDA Challenge: https://precision.fda.gov/challenges/10

Higher accuracy translates to lower false positive base calls



	SBB	NovaSeq
■ False Positive InDel Calls	3323	7264
■ False Positive SNV Calls	2218	5197

Even lower error rate than original emulsion PCR method

Reaches 0.1% “true variation level” between the sample and reference

SBB Over 2x lower false positives than SBS

NovaSeq data from: https://precision.fda.gov/challenges/10

PacBio confidential 7

Sequencing by binding (SBB)

1. Technology
2. Benchmark
3. Application

Developing an innovative platform to house SBB

Novel and state-of-the-art inventions

4 core focus areas

140+ patents pending

~50 patents allowed/issued

Proprietary chemistry

Consumables design

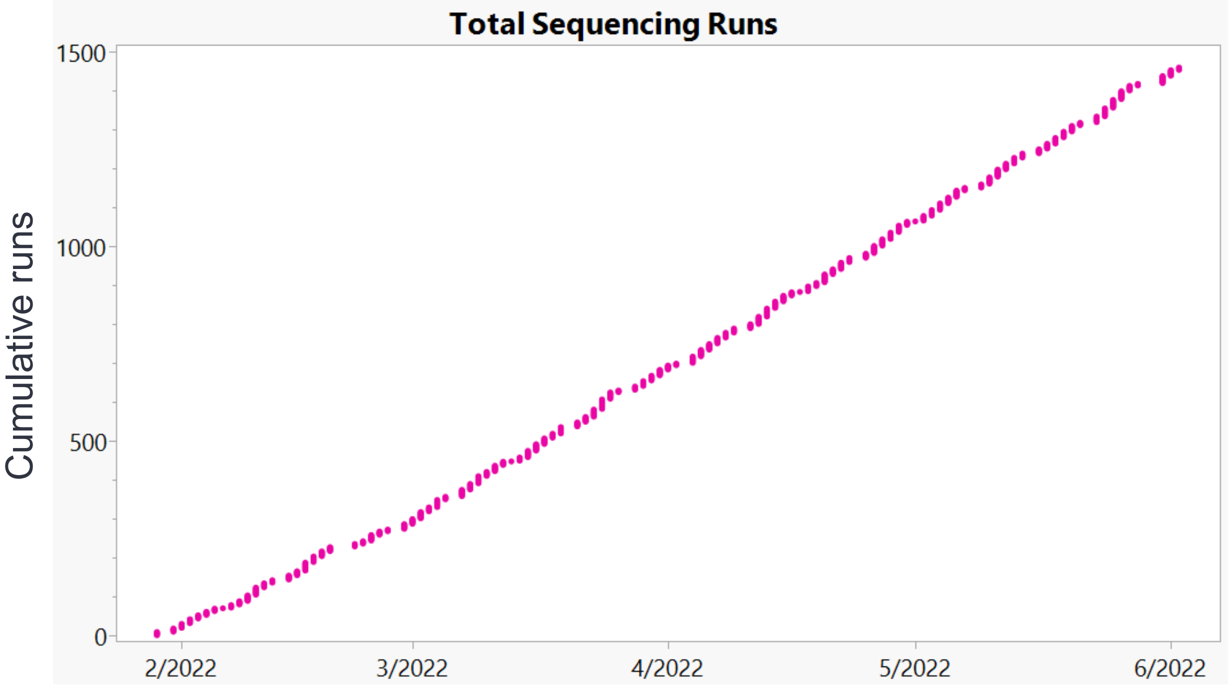
Instrument engineering

Software algorithms

PacBio

Sequencing around the clock

>1500 runs completed in 2022 alone



Success: 70% 89% 92% 92% 95%

PacBio

Breakthrough short-read sequencing

Key design principles and goals

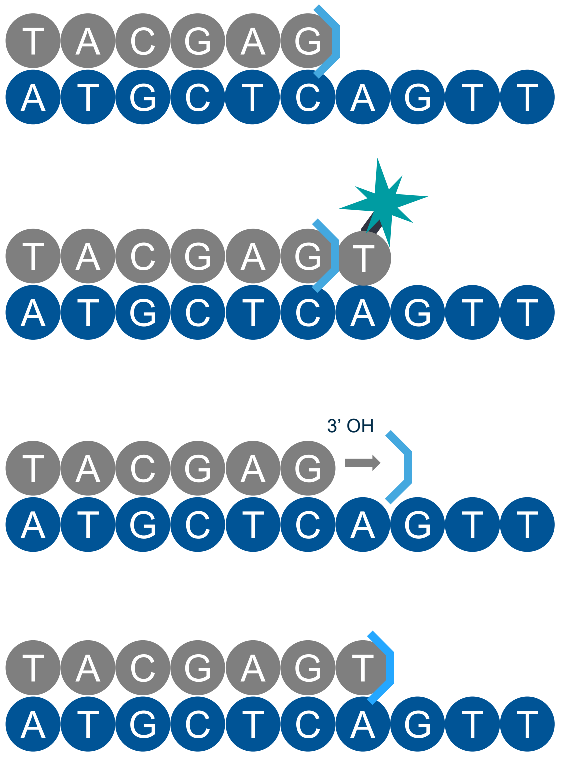
- Mid- to high-throughput NGS platform
- Optical and mechanical innovations
- Scalable, flexible, and cost-optimized
- Unparalleled accuracy from SBB

SBB is fundamentally designed to maximize accuracy

SBB chemistry separates *interrogation* and *incorporation* steps

↓

Multiple optimization points increase accuracy and flexibility



Blocked 3’ end

Interrogate
Flow nucleotides, image, wash

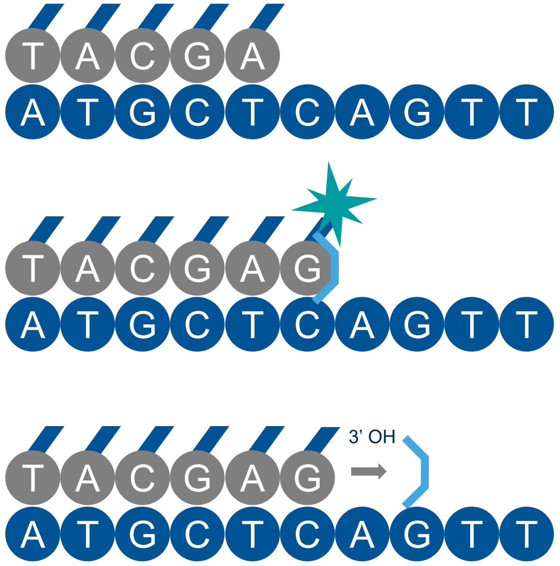
Activate
Remove 3’ RT

Incorporate
Flow blocked nucleotides

SBB advantages: Incorporates native nucleotides, produces unmodified DNA

No base modifications, no molecular scarring

Sequencing by synthesis (SBS)

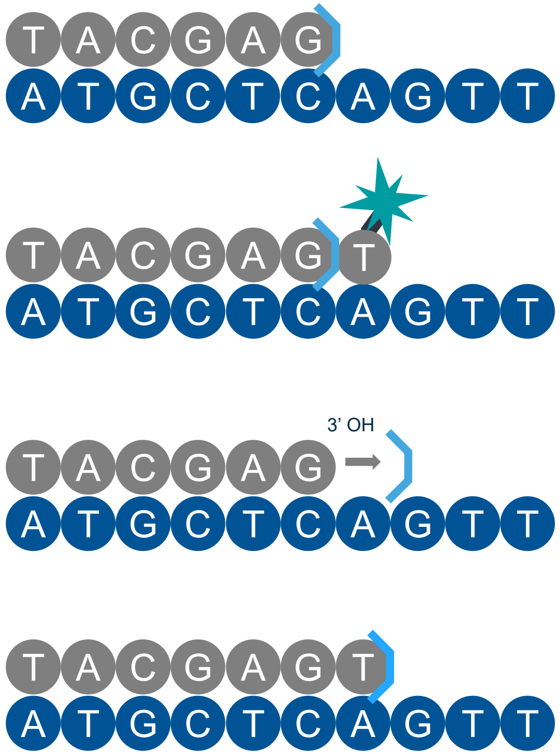


Unblocked 3' end

Incorporate/
Interrogate

Cleavage

Sequencing by binding (SBB)



Blocked 3' end

Interrogate

Activate

Incorporate

Benefits of SBB over traditional short read NGS

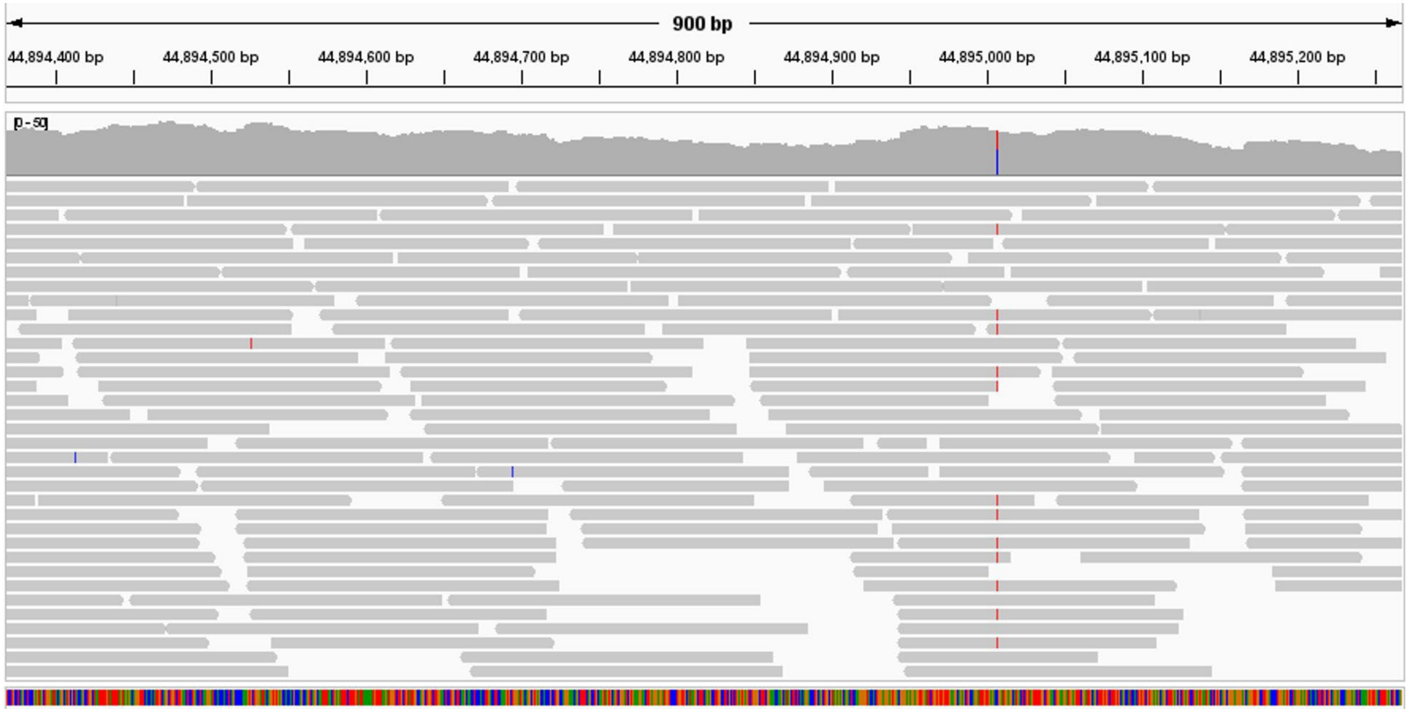
>90% bases at Q40+

Low duplications rate

No index hopping

Sequence through difficult / repetitive regions

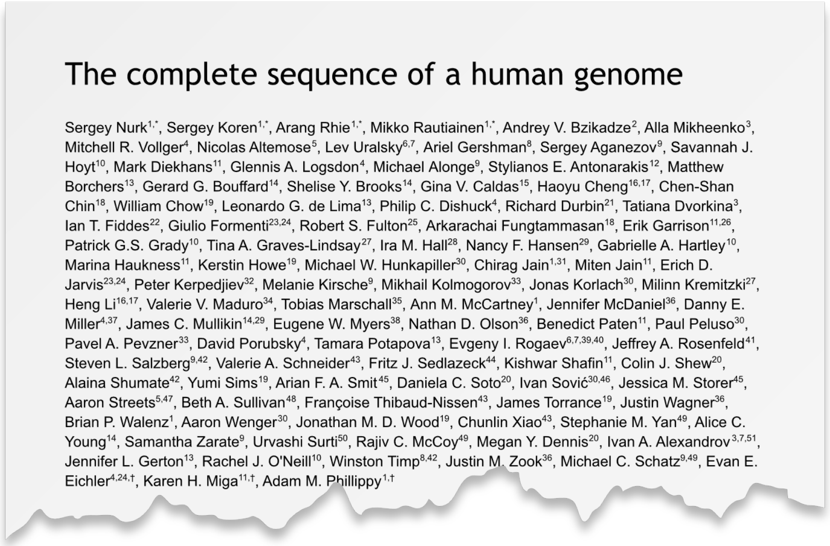
SBB offers near "perfect" reads



HG002; chr19:44,894,368-44,895,266

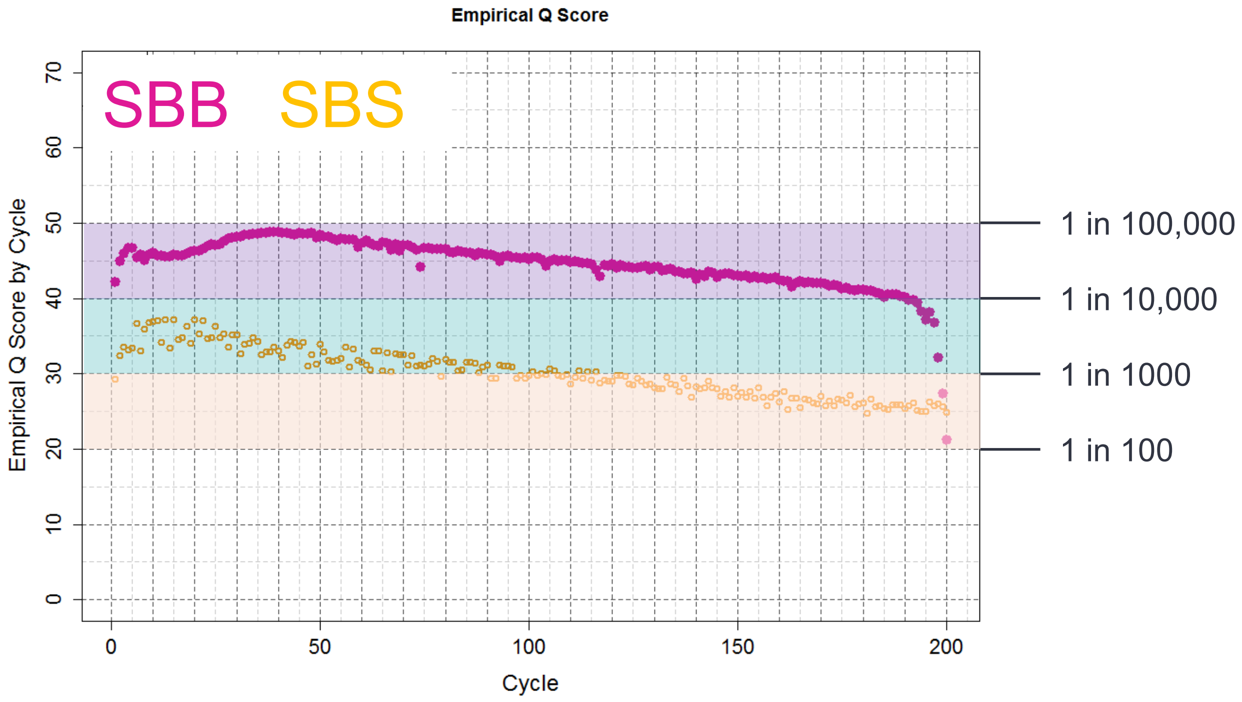
SBB offers best-in-class accuracy

SBB error rates ~15× lower at any given cycle, between 1:10,000 to 1:100,000



The complete sequence of a human genome

Sergey Nurk[1,*], Sergey Koren[1,*], Arang Rhie[1,*], Mikko Rautiainen[1,*], Andrey V. Bzikadze[2], Alla Mikheenko[3], Mitchell R. Vollger[4], Nicolas Altemose[5], Lev Uralsky[6,7], Ariel Gershman[8], Sergey Aganezov[9], Savannah J. Hoyt[10], Mark Diekhans[11], Glennis A. Logsdon[4], Michael Alonge[9], Stylianos E. Antonarakis[12], Matthew Borchers[13], Gerard G. Bouffard[14], Shelise Y. Brooks[14], Gina V. Caldas[15], Haoyu Cheng[16,17], Chen-Shan Chin[18], William Chow[19], Leonardo G. de Lima[13], Philip C. Dishuck[4], Richard Durbin[21], Tatiana Dvorkina[3], Ian T. Fiddes[22], Giulio Formenti[23,24], Robert S. Fulton[25], Arkarachai Fungtammasan[18], Erik Garrison[11,26], Patrick G.S. Grady[10], Tina A. Graves-Lindsay[27], Ira M. Hall[28], Nancy F. Hansen[29], Gabrielle A. Hartley[10], Marina Haukness[11], Kerstin Howe[19], Michael W. Hunkapiller[30], Chirag Jain[1,31], Miten Jain[11], Erich D. Jarvis[23,24], Peter Kerpedjiev[32], Melanie Kirsche[9], Mikhail Kolmogorov[33], Jonas Korlach[30], Milinn Kremitzki[27], Heng Li[16,17], Valerie V. Maduro[34], Tobias Marschall[35], Ann M. McCartney[1], Jennifer McDaniel[36], Danny E. Miller[4,37], James C. Mullikin[14,29], Eugene W. Myers[38], Nathan D. Olson[36], Benedict Paten[11], Paul Peluso[30], Pavel A. Pevzner[33], David Porubsky[4], Tamara Potapova[13], Evgeny I. Rogaev[6,7,39,40], Jeffrey A. Rosenfeld[41], Steven L. Salzberg[9,42], Valerie A. Schneider[43], Fritz J. Sedlazeck[44], Kishwar Shafin[11], Colin J. Shew[20], Alaina Shumate[42], Yumi Sims[19], Arian F. A. Smit[45], Daniela C. Soto[20], Ivan Sović[30,46], Jessica M. Storer[45], Aaron Streets[5,47], Beth A. Sullivan[48], Françoise Thibaud-Nissen[43], James Torrance[19], Justin Wagner[36], Brian P. Walenz[1], Aaron Wenger[30], Jonathan M. D. Wood[19], Chunlin Xiao[43], Stephanie M. Yan[49], Alice C. Young[14], Samantha Zarate[9], Urvashi Surti[50], Rajiv C. McCoy[49], Megan Y. Dennis[20], Ivan A. Alexandrov[3,7,51], Jennifer L. Gerton[13], Rachel J. O'Neill[10], Winston Timp[8,42], Justin M. Zook[36], Michael C. Schatz[9,49], Evan E. Eichler[4,24,†], Karen H. Miga[11,†], Adam M. Phillippy[1,†]



Chm13 sequencing. Empirical Q score evaluated on 99% of bases for each cycle.

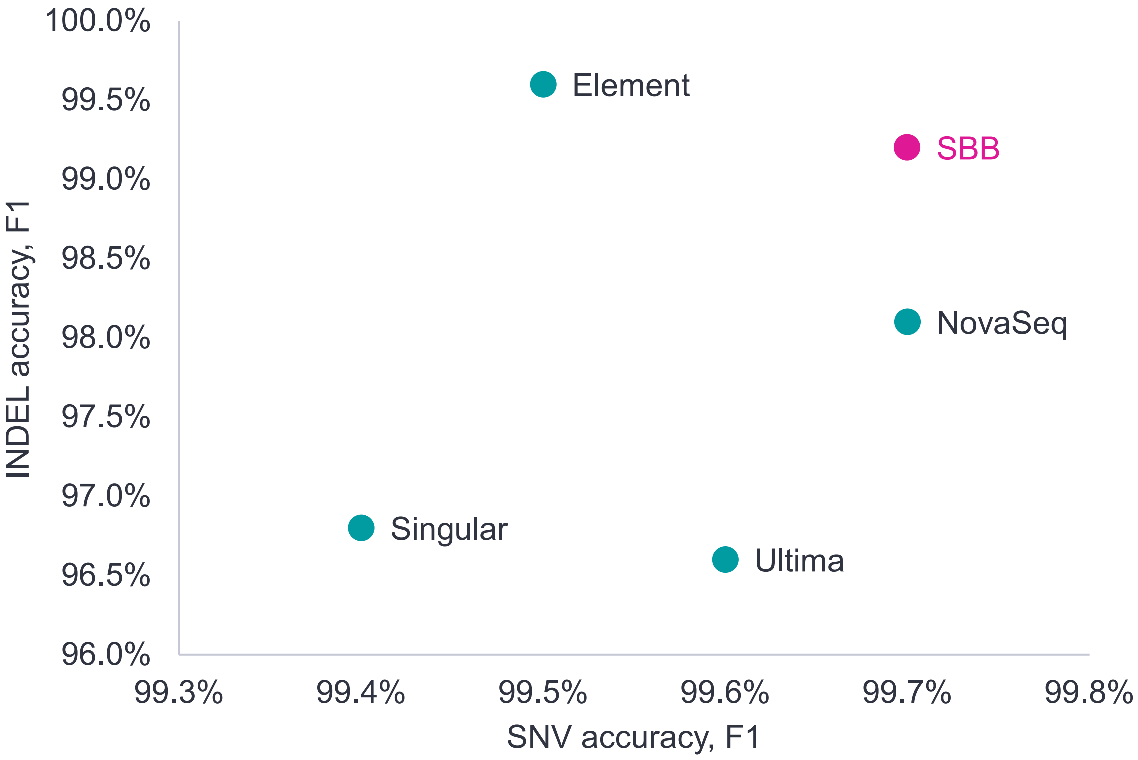
SBB vs SBS empirical vs reported per-base Q score for CHM13

Uncalibrated SBB Q score correlates well with observed errors

Chm13 sequencing. Empirical Q score evaluated on 99% of bases for each cycle.

Excellent variant calling performance for SBB

		SBB	**NovaSeq**	**Element**	**Ultima**	**Singular**
SNV	Recall	99.6%	99.9%	99.1%	99.6%	99.2%
	Prec.	99.7%	99.5%	99.8%	99.6%	99.7%
	F1	**99.7%**	**99.7%**	**99.5%**	**99.6%**	**99.4%**
INDEL	Recall	98.9%	97.9%	99.3%	96.4%	96.4%
	Prec.	99.4%	98.4%	99.8%	96.8%	97.1%
	F1	**99.2%**	**98.1%**	**99.6%**	**96.6%**	**96.8%**



- Expecting further improvements through variant caller training

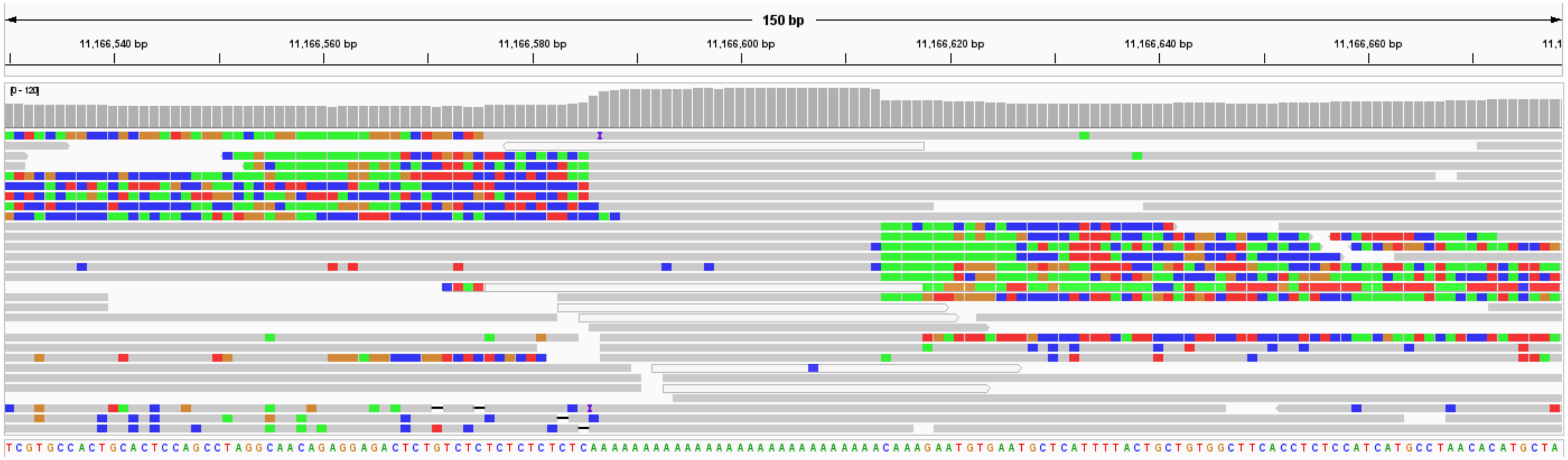
SBB / NVSQ: HG002 GIAB benchmark v4.2, Sentieon Haplotyper v20211204
Element: https://go.elementbiosciences.com/access-app-note-roche-human-microbial
Ultima: https://doi.org/10.1101/2022.05.29.493900 HG002 40x coverage
Singular: G4 Human Genome Sequencing Technical Report 1.0. NA12878 40x coverage

What does unprecedented accuracy look like?

Low-complexity region (28 As in the reference)

SBS

reference



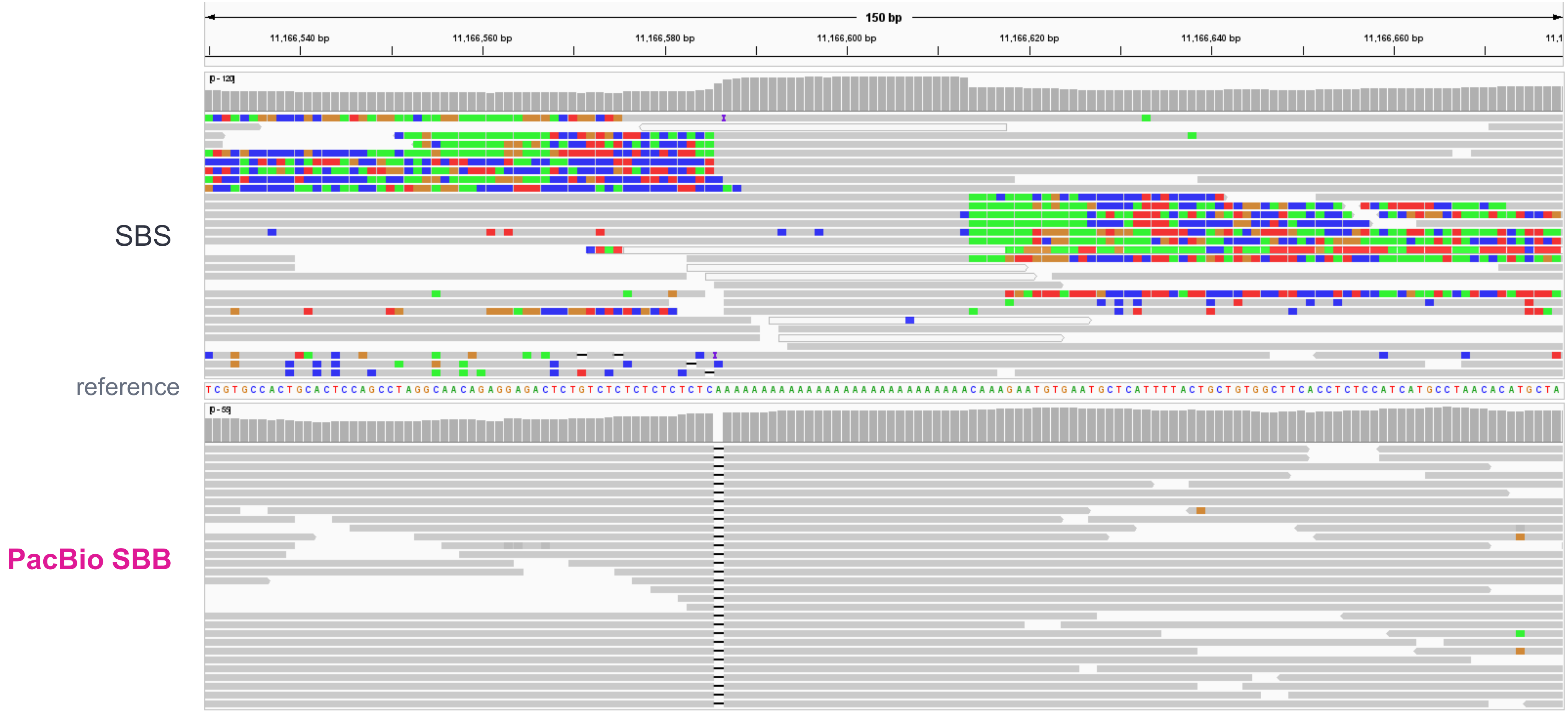
No high-quality reads

No variants called

CHM13; chr5:11,166,530-11,166,679. PacBio SBB 1x200; HiSeq 2x250

What does unprecedented accuracy look like?

SBB cleanly sequences through 27 bp poly A (28 bp in the reference)



No high-quality reads

No variants called

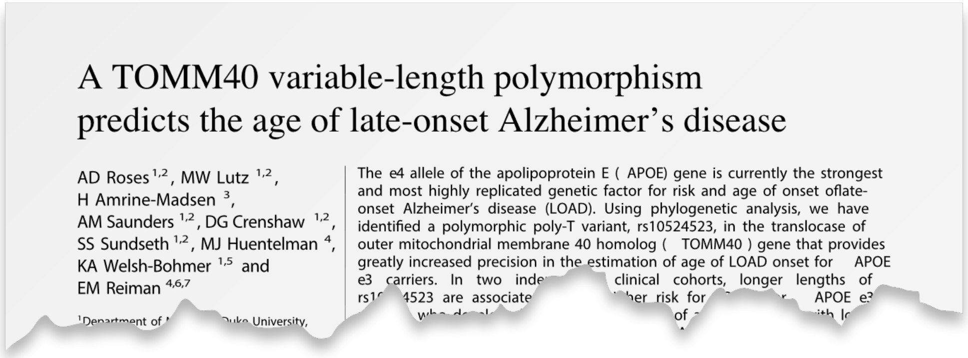
Full resolution

1 bp deletion (27 As)

PacBio

CHM13; chr5:11,166,530-11,166,679. PacBio SBB 1x200; HiSeq 2x250

Example of a ‘difficult’ region – *TOMM40*

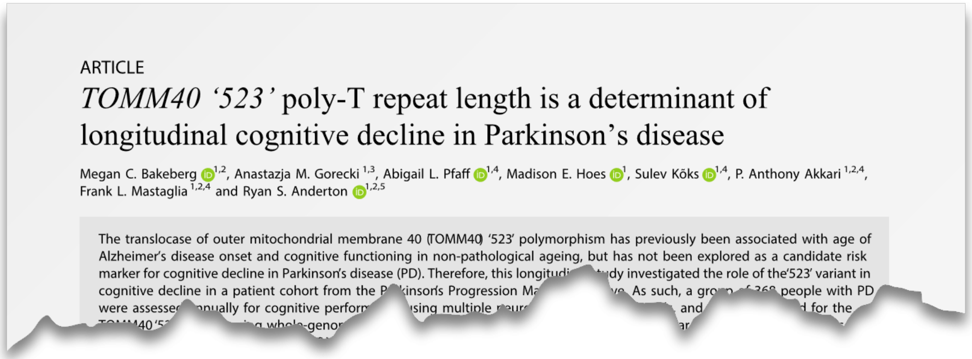


A TOMM40 variable-length polymorphism predicts the age of late-onset Alzheimer’s disease

AD Roses[1,2], MW Lutz[1,2],
H Amrine-Madsen[3],
AM Saunders[1,2], DG Crenshaw[1,2],
SS Sundseth[1,2], MJ Huentelman[4],
KA Welsh-Bohmer[1,5] and
EM Reiman[4,6,7]

[1]Department of … Duke University,

The e4 allele of the apolipoprotein E (APOE) gene is currently the strongest and most highly replicated genetic factor for risk and age of onset oflate-onset Alzheimer’s disease (LOAD). Using phylogenetic analysis, we have identified a polymorphic poly-T variant, rs10524523, in the translocase of outer mitochondrial membrane 40 homolog (TOMM40) gene that provides greatly increased precision in the estimation of age of LOAD onset for APOE e3 carriers. In two inde… clinical cohorts, longer lengths of rs1…4523 are associate… …her risk for … APOE e3 … who … of … ith l…



ARTICLE

TOMM40 ‘523’ poly-T repeat length is a determinant of longitudinal cognitive decline in Parkinson’s disease

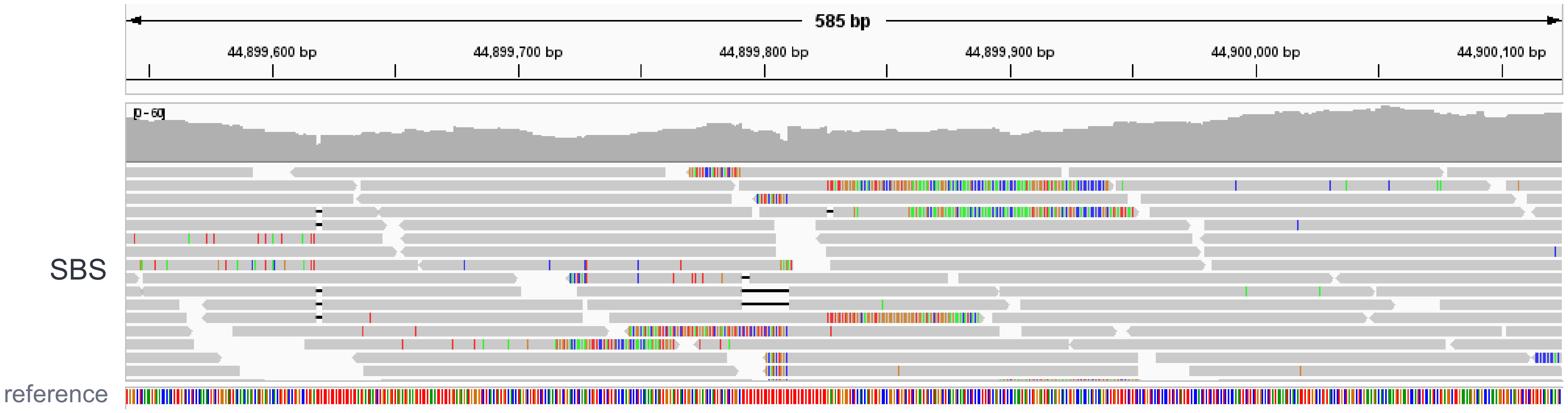
Megan C. Bakeberg[1,2], Anastazja M. Gorecki[1,3], Abigail L. Pfaff[1,4], Madison E. Hoes[1], Sulev Kõks[1,4], P. Anthony Akkari[1,2,4], Frank L. Mastaglia[1,2,4] and Ryan S. Anderton[1,2,5]

The translocase of outer mitochondrial membrane 40 (TOMM40) ‘523’ polymorphism has previously been associated with age of Alzheimer’s disease onset and cognitive functioning in non-pathological ageing, but has not been explored as a candidate risk marker for cognitive decline in Parkinson’s disease (PD). Therefore, this longitudi… …dy investigated the role of the‘523’ variant in cognitive decline in a patient cohort from the P…kinson’s Progression Ma… …ve. As such, a gro… people with PD were assesse… …nually for cognitive perform… …sing multiple neur… and … …d for the TOMM40 ‘5… …whole-genor… …ar…

“Assay development for the TOMM40 ‘523’ variant is generally considered to be difficult, as poly-T variants are challenging to sequence.” [1]

Adapted from Larsen *et al.* (2017) https://alz-journals.onlinelibrary.wiley.com/doi/10.1016/j.jalz.2017.01.017; https://www.nature.com/articles/s41531-021-00200-y; https://pubmed.ncbi.nlm.nih.gov/20029386/ [1]https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7973542/

Example of a ‘difficult’ region – *TOMM40*



No high-quality reads

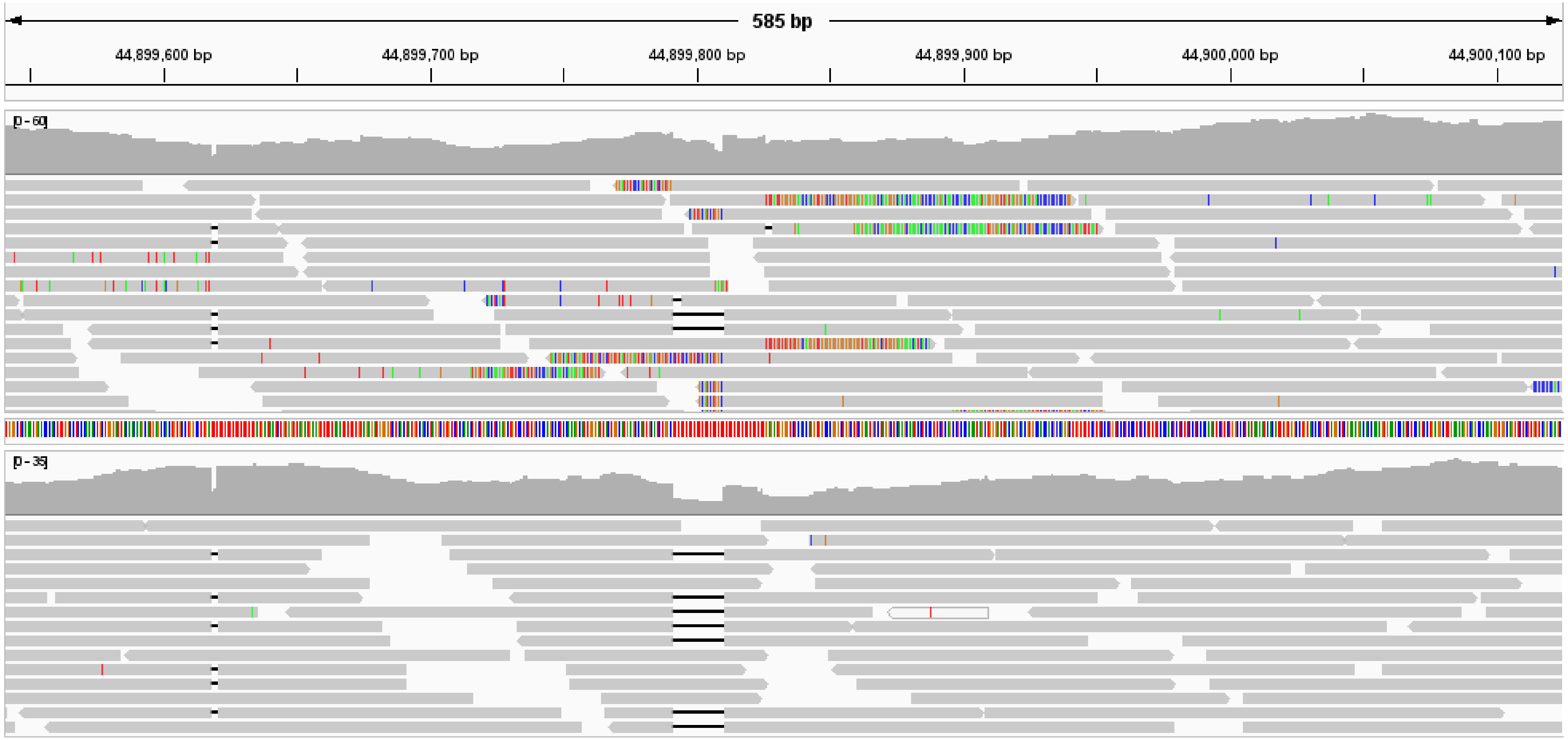
poly-T repeat unresolved

poly-T repeat

HG002; chr19:44,899,541-44,900,124; PacBio SBB 1x200 bp; NovaSeq 2x151 bp

Example of a ‘difficult’ region – *TOMM40*

SBB cleanly sequences through poly-T repeat locus



SBS

No high-quality reads

poly-T repeat unresolved

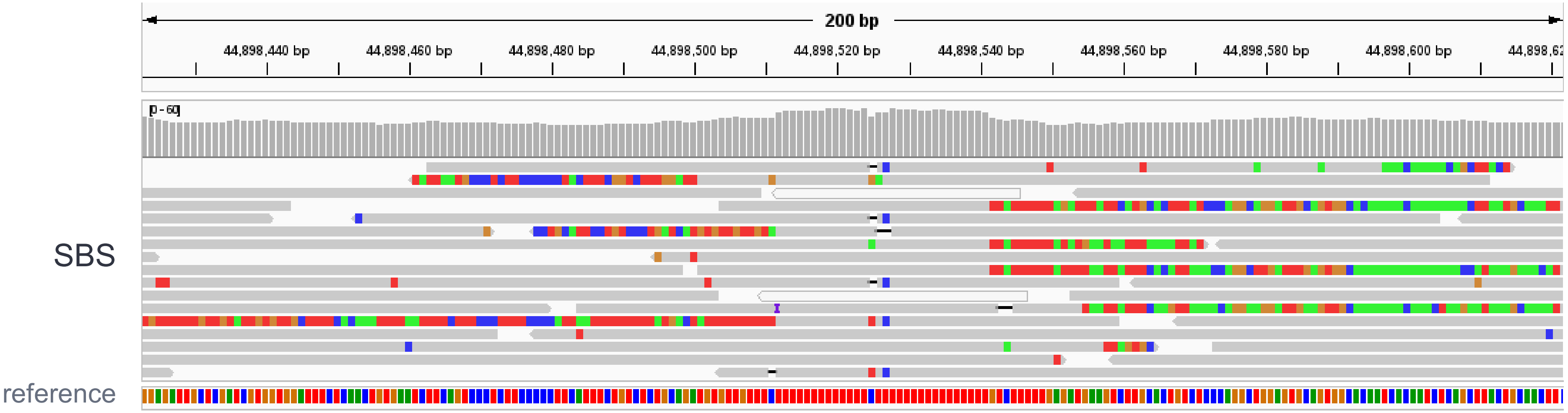
PacBio SBB

poly-T repeat fully resolved: T_{16} / T_{35}

poly-T repeat

HG002; chr19:44,899,541-44,900,124; PacBio SBB 1x200 bp; NovaSeq 2x151 bp

Another similar region nearby

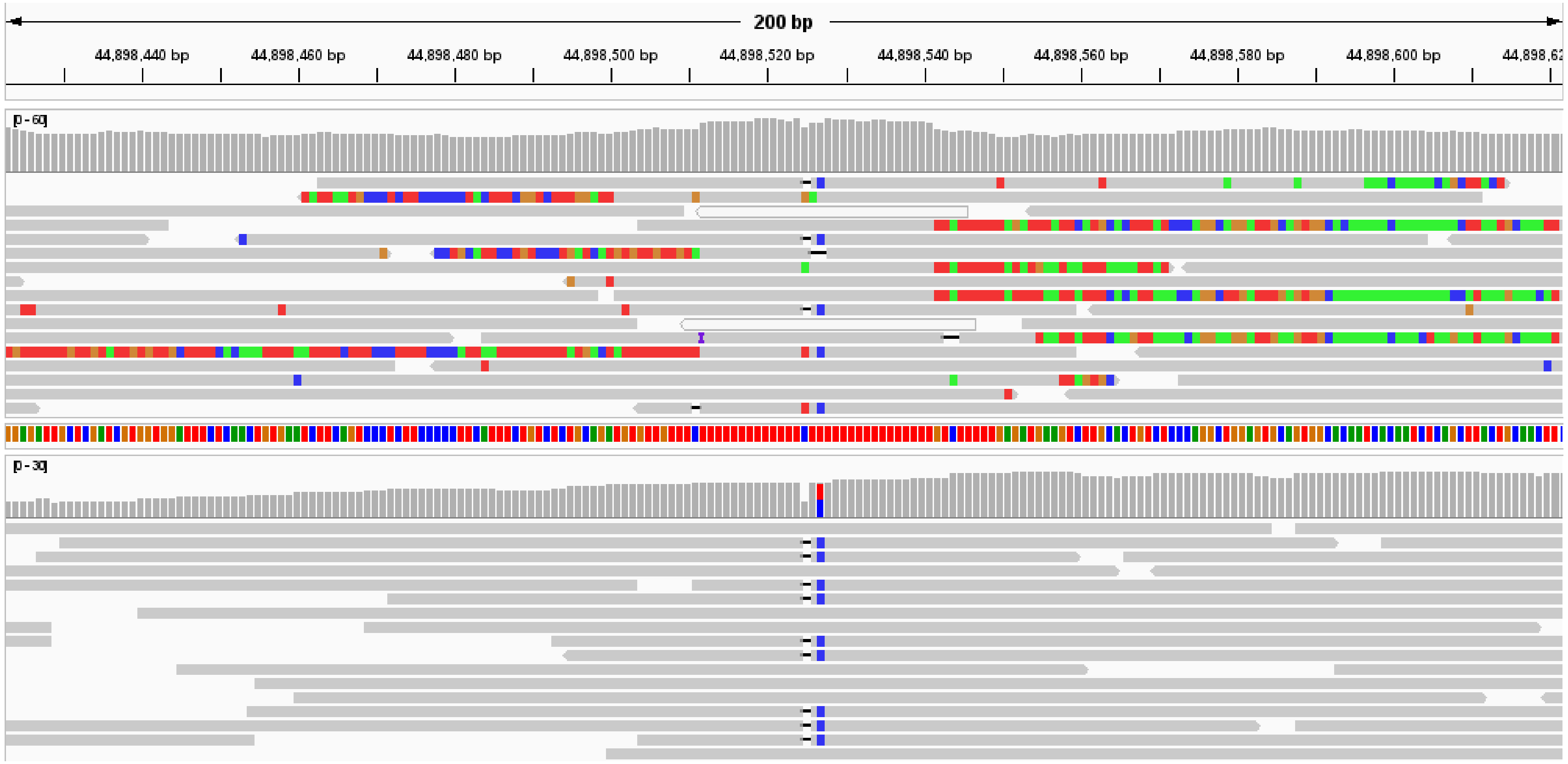


No high-quality reads

No variants called

HG002; chr19:44,898,425-44,898,624; PacBio SBB 1x200 bp; NovaSeq 2x151 bp

Another similar region nearby



SBS

reference

PacBio SBB

No high-quality reads

No variants called

Two het variants called

Full resolution & phasing:

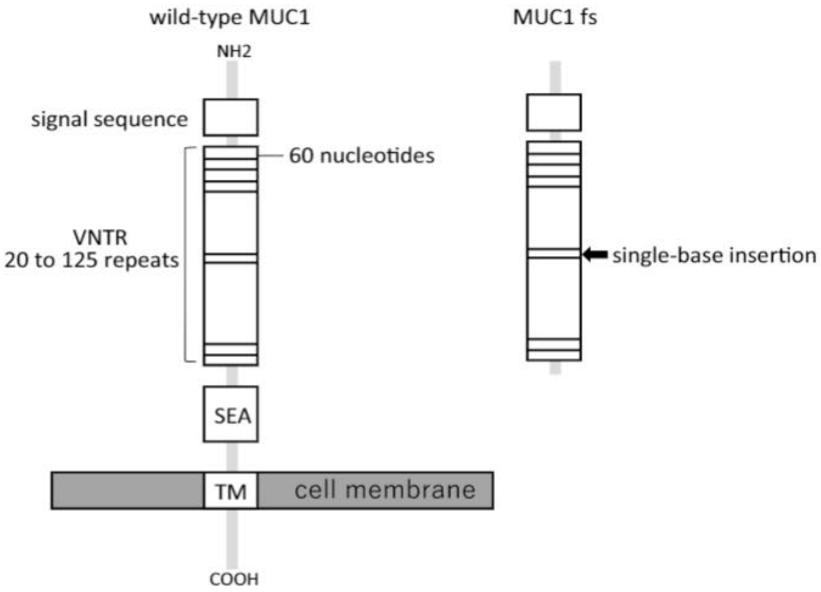
`TTTTTTTTTTTTTCTTTTTTTTTTTTTTTT`
`TTTTTTTTTTTTT-TCTTTTTTTTTTTTTT`

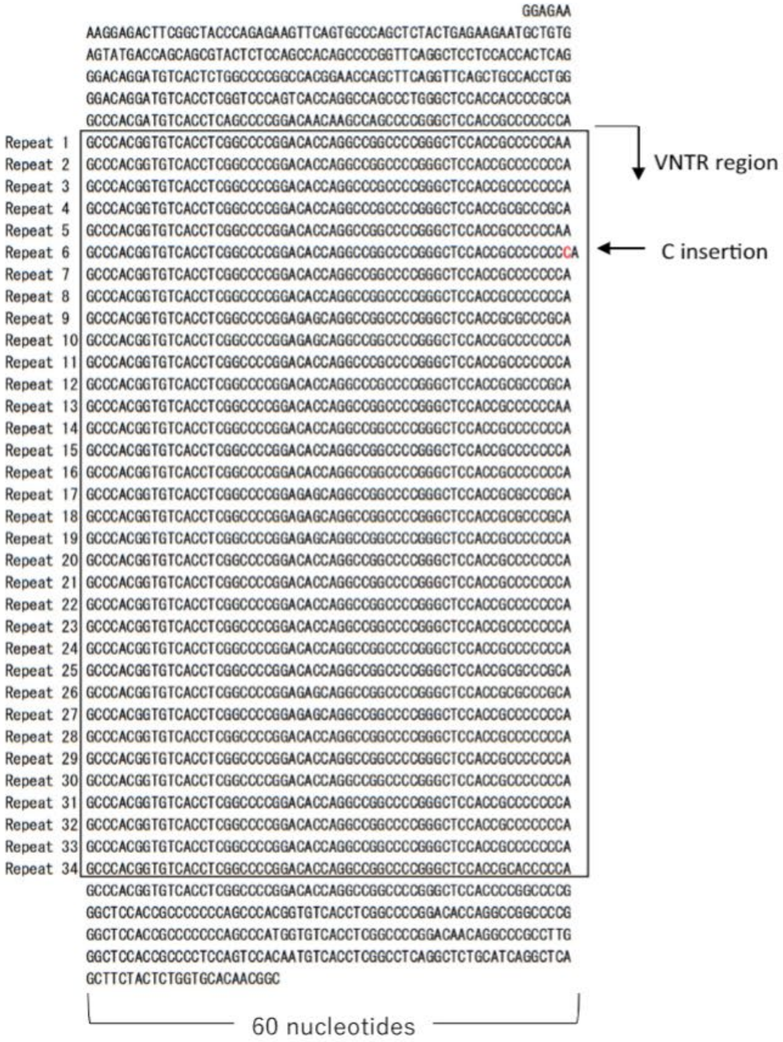
PacBio

HG002; chr19:44,898,425-44,898,624; PacBio SBB 1x200 bp; NovaSeq 2x151 bp

SBB correctly sequences through lengthy C/G repeats

Example Mucin 1 (*MUC1*) kidney disease



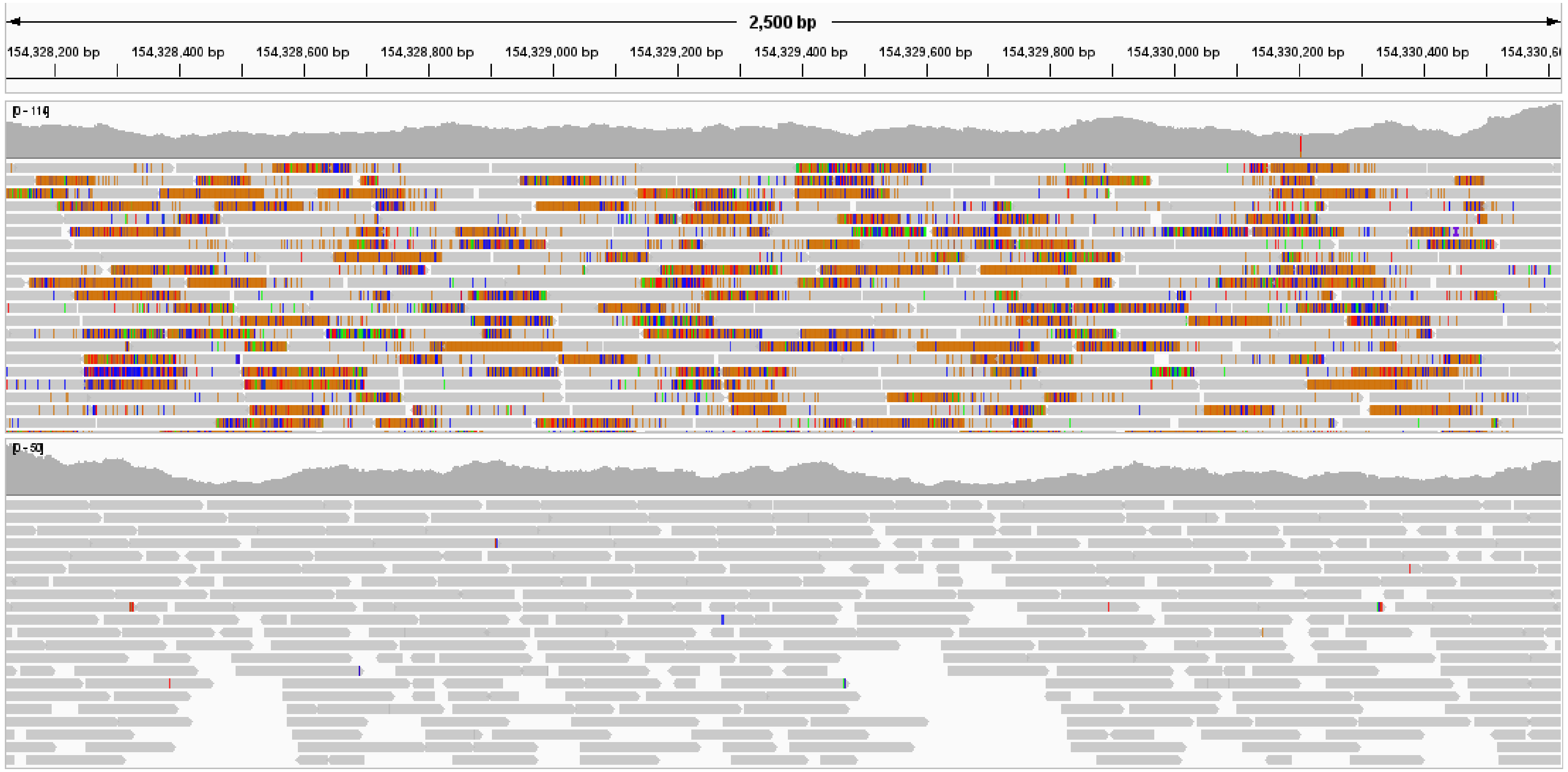


```
                                                                  GGAGAA
           AAGGAGACTTCGGCTACCCAGAGAAGTTCAGTGCCCAGCTCTACTGAGAAGAATGCTGTG
           AGTATGACCAGCAGCGTACTCTCCAGCCACAGCCCCGGTTCAGGCTCCTCCACCACTCAG
           GGACAGGATGTCACTCTGGCCCCGGCCACGGAACCAGCTTCAGGTTCAGCTGCCACCTGG
           GGACAGGATGTCACCTCGGTCCCAGTCACCAGGCCAGCCCTGGGCTCCACCACCCCGCCA
           GCCCACGATGTCACCTCAGCCCCGGACAACAAGCCAGCCCCGGGCTCCACCGCCCCCCCA
Repeat  1  GCCCACGGTGTCACCTCGGCCCCGGACACCAGGCCGGCCCCGGGCTCCACCGCCCCCCAA
Repeat  2  GCCCACGGTGTCACCTCGGCCCCGGACACCAGGCCGGCCCCGGGCTCCACCGCCCCCCCA
Repeat  3  GCCCACGGTGTCACCTCGGCCCCGGACACCAGGCCCGCCCCGGGCTCCACCGCCCCCCCA
Repeat  4  GCCCACGGTGTCACCTCGGCCCCGGACACCAGGCCCGCCCCGGGCTCCACCGCGCCCGCA
Repeat  5  GCCCACGGTGTCACCTCGGCCCCGGACACCAGGCCGGCCCCGGGCTCCACCGCCCCCCAA
Repeat  6  GCCCACGGTGTCACCTCGGCCCCGGACACCAGGCCGGCCCCGGGCTCCACCGCCCCCCCCA
Repeat  7  GCCCACGGTGTCACCTCGGCCCCGGACACCAGGCCGGCCCCGGGCTCCACCGCCCCCCCA
Repeat  8  GCCCACGGTGTCACCTCGGCCCCGGACACCAGGCCGGCCCCGGGCTCCACCGCCCCCCCA
Repeat  9  GCCCACGGTGTCACCTCGGCCCCGGAGAGCAGGCCGGCCCCGGGCTCCACCGCGCCCGCA
Repeat 10  GCCCACGGTGTCACCTCGGCCCCGGAGAGCAGGCCGGCCCCGGGCTCCACCGCCCCCCCA
Repeat 11  GCCCACGGTGTCACCTCGGCCCCGGACACCAGGCCCGCCCCGGGCTCCACCGCCCCCCCA
Repeat 12  GCCCACGGTGTCACCTCGGCCCCGGACACCAGGCCCGCCCCGGGCTCCACCGCGCCCGCA
Repeat 13  GCCCACGGTGTCACCTCGGCCCCGGACACCAGGCCGGCCCCGGGCTCCACCGCCCCCCAA
Repeat 14  GCCCACGGTGTCACCTCGGCCCCGGACACCAGGCCGGCCCCGGGCTCCACCGCCCCCCCA
Repeat 15  GCCCACGGTGTCACCTCGGCCCCGGACACCAGGCCGGCCCCGGGCTCCACCGCCCCCCCA
Repeat 16  GCCCACGGTGTCACCTCGGCCCCGGACACCAGGCCGGCCCCGGGCTCCACCGCCCCCCCA
Repeat 17  GCCCACGGTGTCACCTCGGCCCCGGAGAGCAGGCCGGCCCCGGGCTCCACCGCGCCCGCA
Repeat 18  GCCCACGGTGTCACCTCGGCCCCGGAGAGCAGGCCGGCCCCGGGCTCCACCGCGCCCGCA
Repeat 19  GCCCACGGTGTCACCTCGGCCCCGGAGAGCAGGCCGGCCCCGGGCTCCACCGCCCCCCCA
Repeat 20  GCCCACGGTGTCACCTCGGCCCCGGACACCAGGCCGGCCCCGGGCTCCACCGCCCCCCCA
Repeat 21  GCCCACGGTGTCACCTCGGCCCCGGACACCAGGCCGGCCCCGGGCTCCACCGCCCCCCCA
Repeat 22  GCCCACGGTGTCACCTCGGCCCCGGACACCAGGCCGGCCCCGGGCTCCACCGCCCCCCCA
Repeat 23  GCCCACGGTGTCACCTCGGCCCCGGACACCAGGCCGGCCCCGGGCTCCACCGCCCCCCCA
Repeat 24  GCCCACGGTGTCACCTCGGCCCCGGACACCAGGCCGGCCCCGGGCTCCACCGCCCCCCCA
Repeat 25  GCCCACGGTGTCACCTCGGCCCCGGACACCAGGCCGGCCCCGGGCTCCACCGCGCCCGCA
Repeat 26  GCCCACGGTGTCACCTCGGCCCCGGAGAGCAGGCCGGCCCCGGGCTCCACCGCGCCCGCA
Repeat 27  GCCCACGGTGTCACCTCGGCCCCGGAGAGCAGGCCGGCCCCGGGCTCCACCGCCCCCCCA
Repeat 28  GCCCACGGTGTCACCTCGGCCCCGGACACCAGGCCGGCCCCGGGCTCCACCGCCCCCCCA
Repeat 29  GCCCACGGTGTCACCTCGGCCCCGGACACCAGGCCGGCCCCGGGCTCCACCGCCCCCCCA
Repeat 30  GCCCACGGTGTCACCTCGGCCCCGGACACCAGGCCGGCCCCGGGCTCCACCGCCCCCCCA
Repeat 31  GCCCACGGTGTCACCTCGGCCCCGGACACCAGGCCGGCCCCGGGCTCCACCGCCCCCCCA
Repeat 32  GCCCACGGTGTCACCTCGGCCCCGGACACCAGGCCGGCCCCGGGCTCCACCGCCCCCCCA
Repeat 33  GCCCACGGTGTCACCTCGGCCCCGGACACCAGGCCGGCCCCGGGCTCCACCGCCCCCCCA
Repeat 34  GCCCACGGTGTCACCTCGGCCCCGGACACCAGGCCGGCCCCGGGCTCCACCGCACCCCCA
           GCCCACGGTGTCACCTCGGCCCCGGACACCAGGCCGGCCCCGGGCTCCACCCCGGCCCCG
           GGCTCCACCGCCCCCCCAGCCCACGGTGTCACCTCGGCCCCGGACACCAGGCCGGCCCCG
           GGCTCCACCGCCCCCCCAGCCCATGGTGTCACCTCGGCCCCGGACAACAGGCCCGCCTTG
           GGCTCCACCGCCCCTCCAGTCCACAATGTCACCTCGGCCTCAGGCTCTGCATCAGGCTCA
           GCTTCTACTCTGGTGCACAACGGC
```

From Okada *et al.* (2022): https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9039432/

Sequencing performance on a healthy control sample

SBB did not observe spurious mutant variants in CHM13



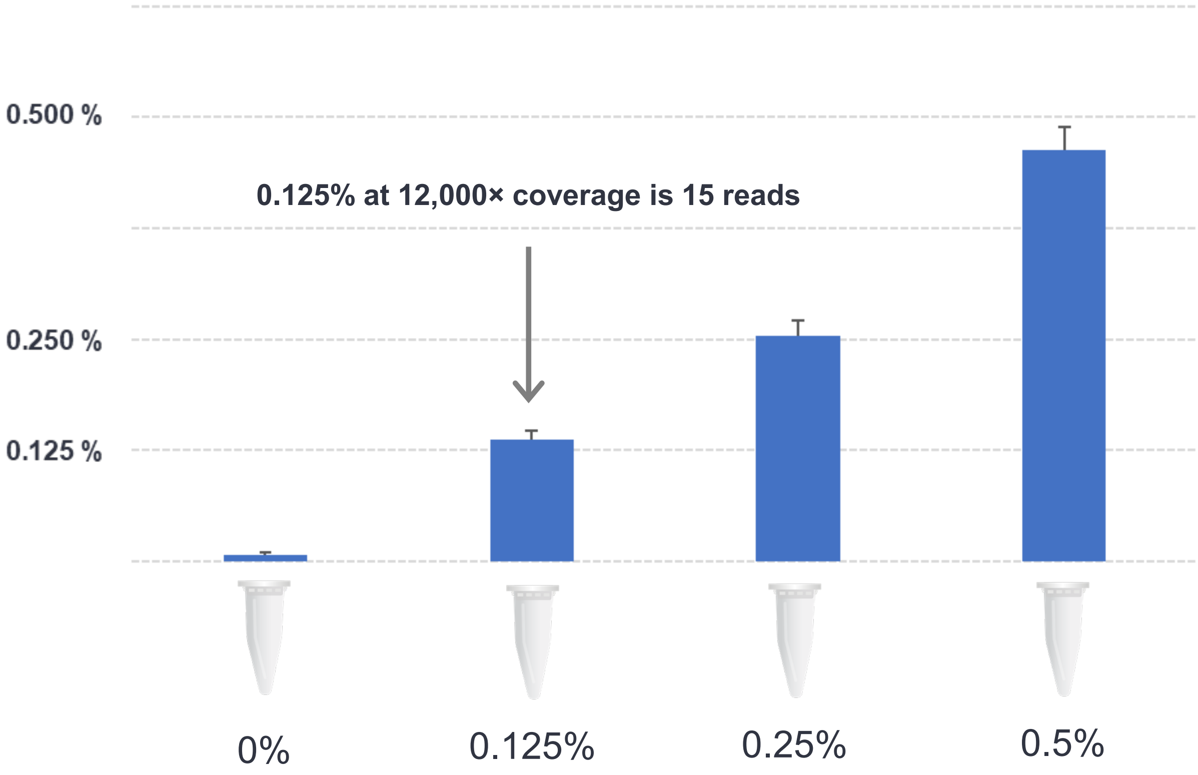
	Observed wildtype CCCCCCC	Observed pathogenic variant CCCCCCCC
SBS	1280	68
PacBio SBB	494	0

CHM13 chr1:154,328,121-154,330,619; PacBio SBB 1x200 bp; HiSeq 1x250 bp

Observed *vs* expected ctDNA at low variant allele frequency

Variant allele percent shows good linearity, even without use of UMIs

High sensitivity and specificity down to 0.125%, even with modest (<12,000×) coverage



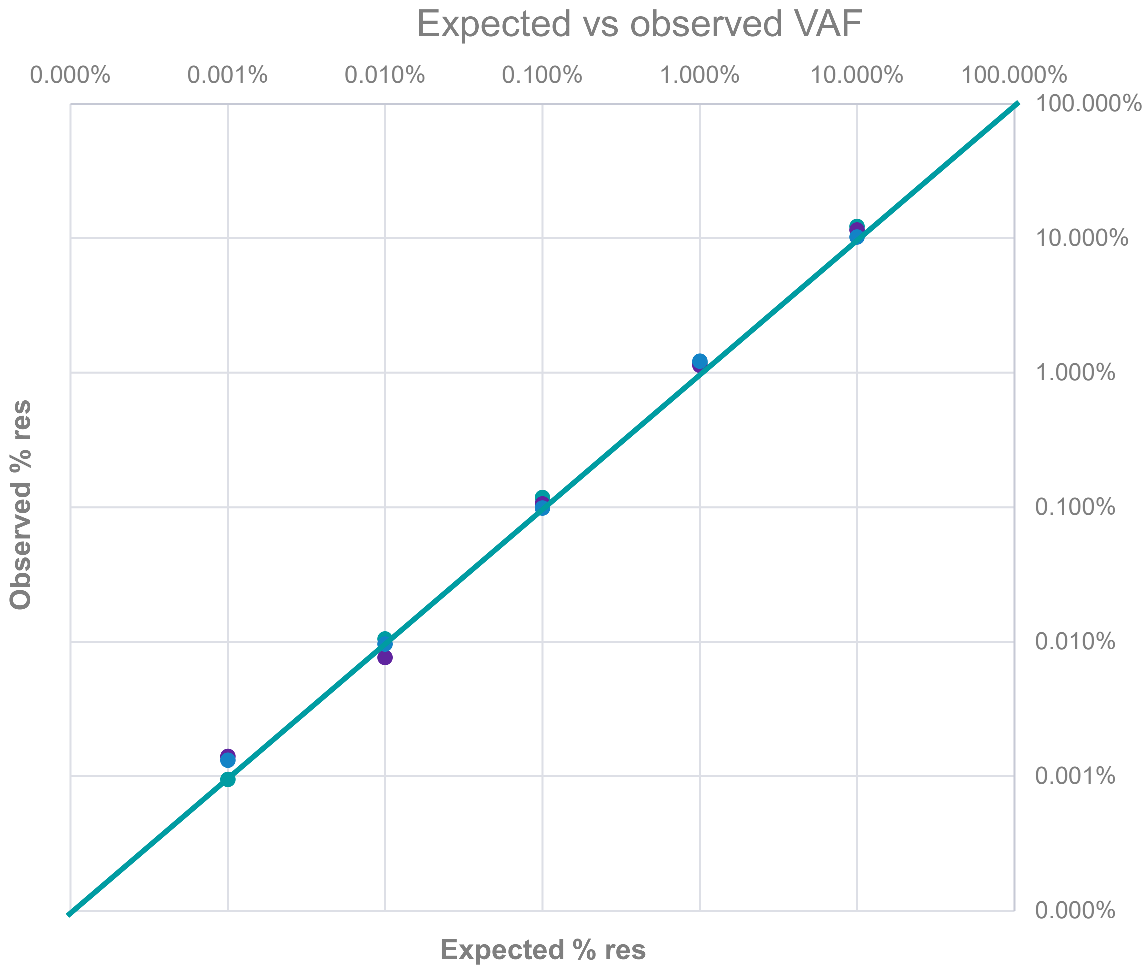
Controls and reference materials
Seraseq® ctDNA Mutation Mix v2

SeraCare control sample

How low can SBB go without UMIs?

Tuberculosis amplicon shows near perfect linearity from 10% to 0.001%

	Res (C) counts	WT (G) counts	Total # counts	Observed % res
10%_rep1	2580549	18479388	21,059,937	12.2534%
10%_rep2	2998687	23062382	26,061,069	11.5064%
10%_rep3	743168	6508219	7,251,387	10.2486%
1%_rep1	76912	6388097	6,465,009	1.1897%
1%_rep2	46152	4025690	4,071,842	1.1334%
1%_rep3	65964	5334716	5,400,680	1.2214%
0.1%_rep1	22836	19283363	19,306,199	0.1183%
0.1%_rep2	8121	7646655	7,654,776	0.1061%
0.1%_rep3	6505	6587930	6,594,435	0.0986%
0.01%_rep1	1164	11078726	11,079,890	0.0105%
0.01%_rep2	197	2579126	2,579,323	0.0076%
0.01%_rep3	1030	10684737	10,685,767	0.0096%
0.001%_rep1	171	18047826	18,047,997	0.0009%
0.001%_rep2	117	8328474	8,328,591	0.0014%
0.001%_rep3	786	59804418	59,805,204	0.0013%



Where to from here for SBB?

Today

- Taking applications for collaboration; run your samples in our lab on SBB
- Visit our suite or www.pacb.com/sbb

Tomorrow

- Thurs, June 9, 8:00–8:30 am: *Advancing NGS accuracy by an order of magnitude*
  Jennifer Stone, PhD, Vice President, Segment Marketing

Late Q3

- Formal external beta commences; more information to be shared at ASHG

1H 23

- On track for platform commercial availability


PacBio