

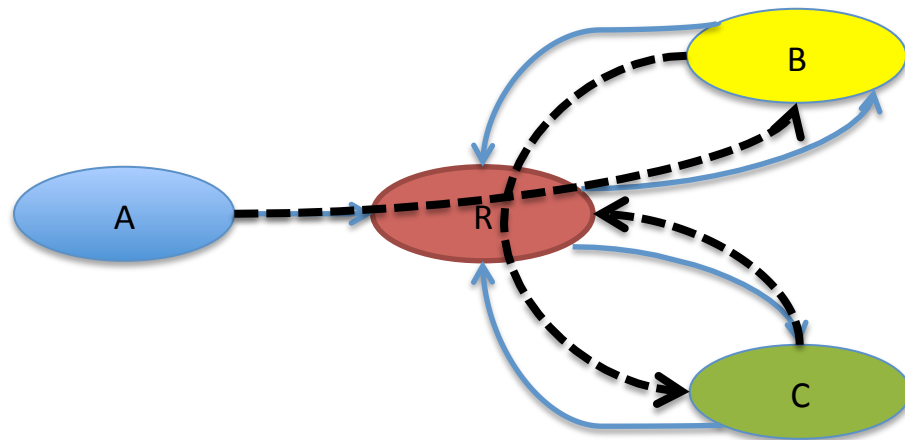
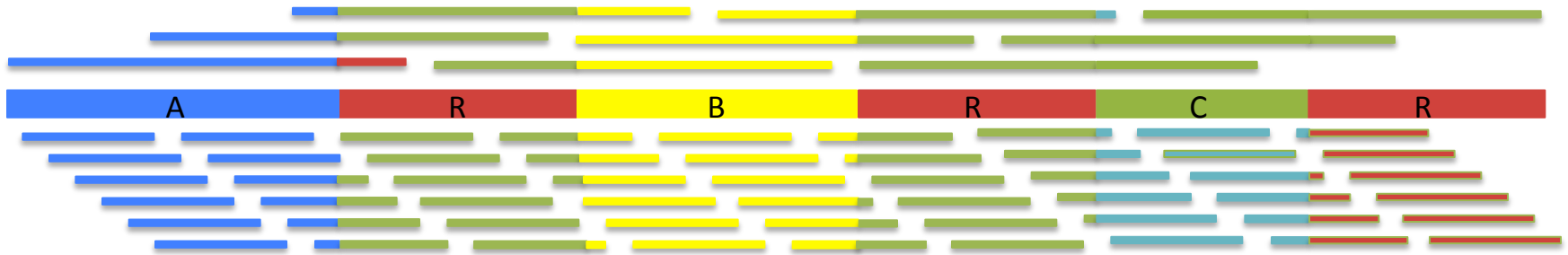


Cold Spring Harbor Laboratory

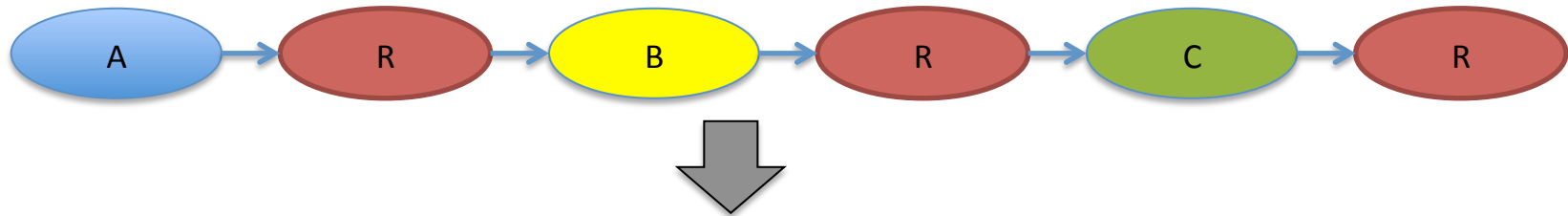
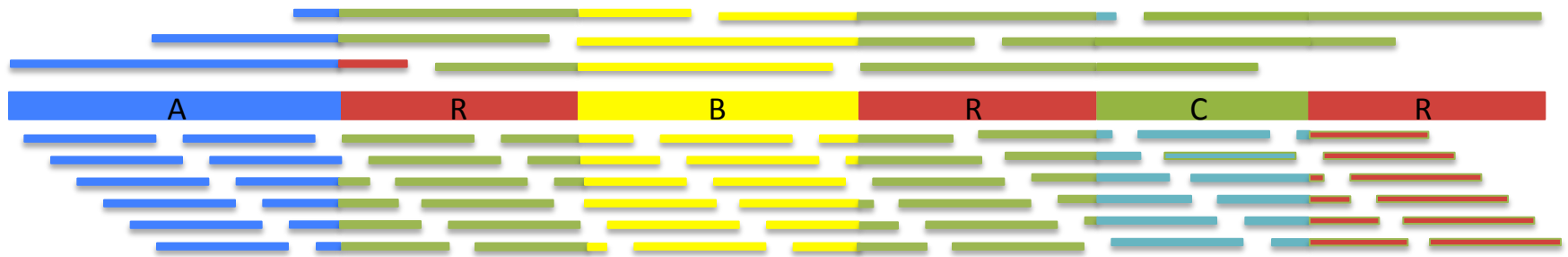
Error Correction and Assembly of Oxford Nanopore Sequencing

James Gurtowski

Assembly Complexity



Assembly Complexity



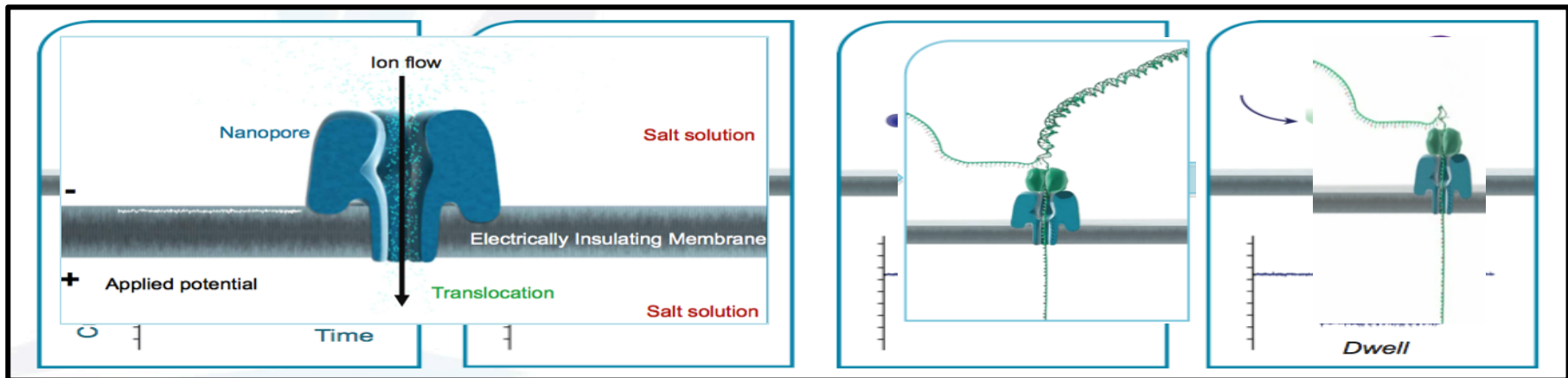
The advantages of SMRT sequencing

Roberts, RJ, Carneiro, MO, Schatz, MC (2013) *Genome Biology*. 14:405

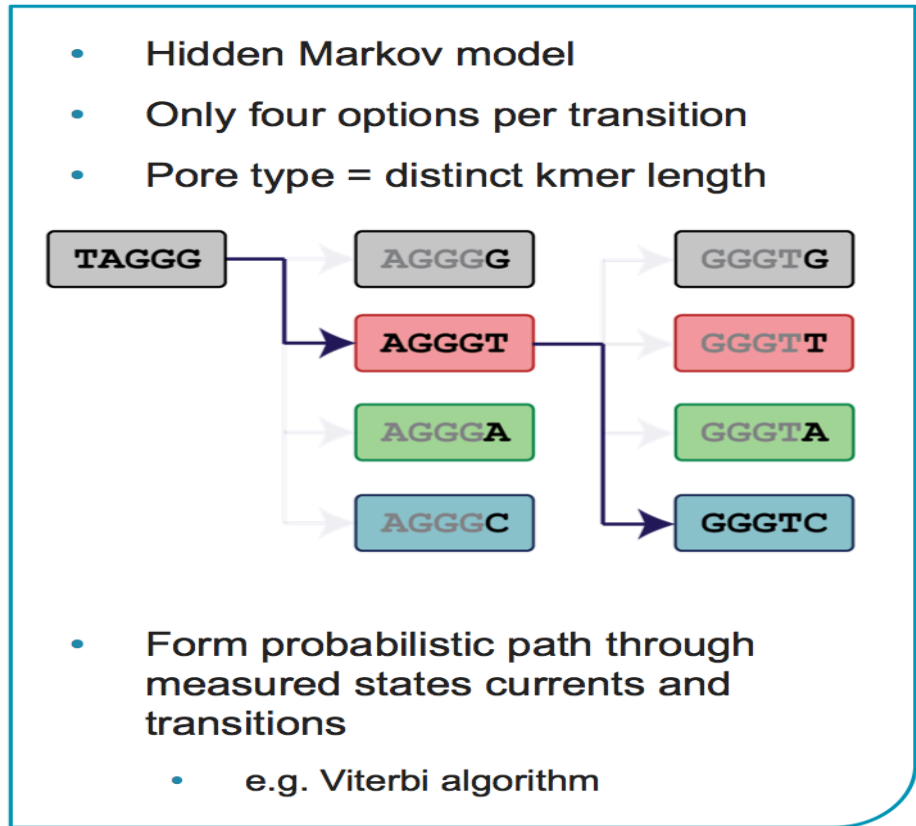
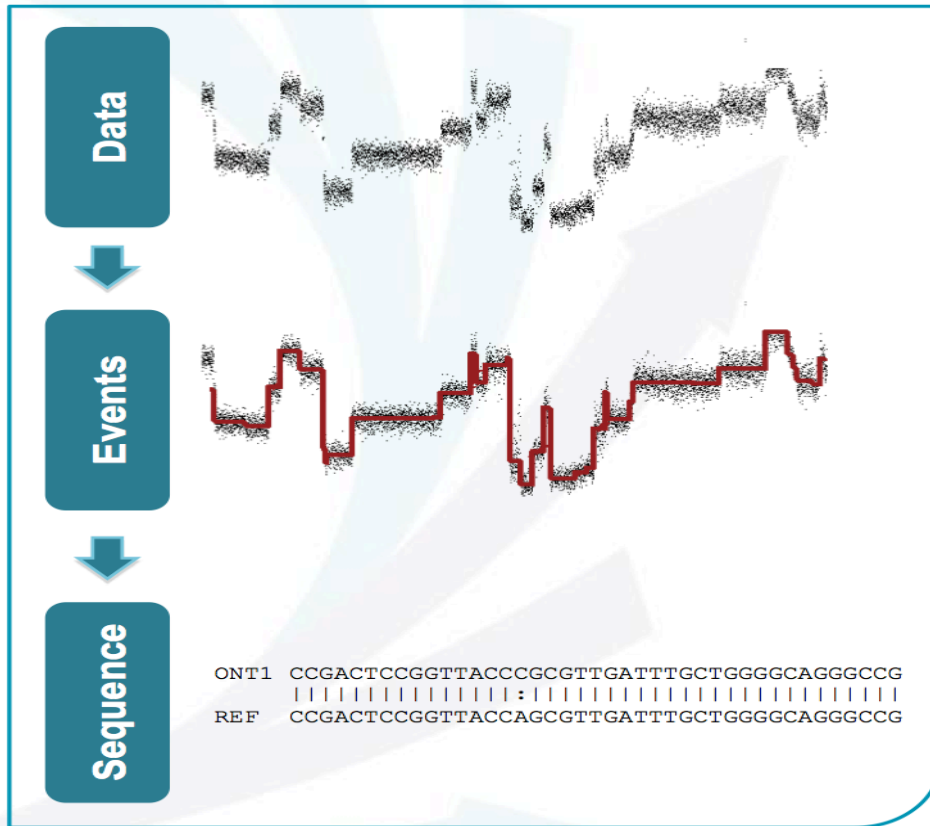
Oxford Nanopore MinION



- Thumb drive sized sequencer powered over USB
- Senses DNA by measuring changes to ion flow
- Reads both DNA Strands (2D)



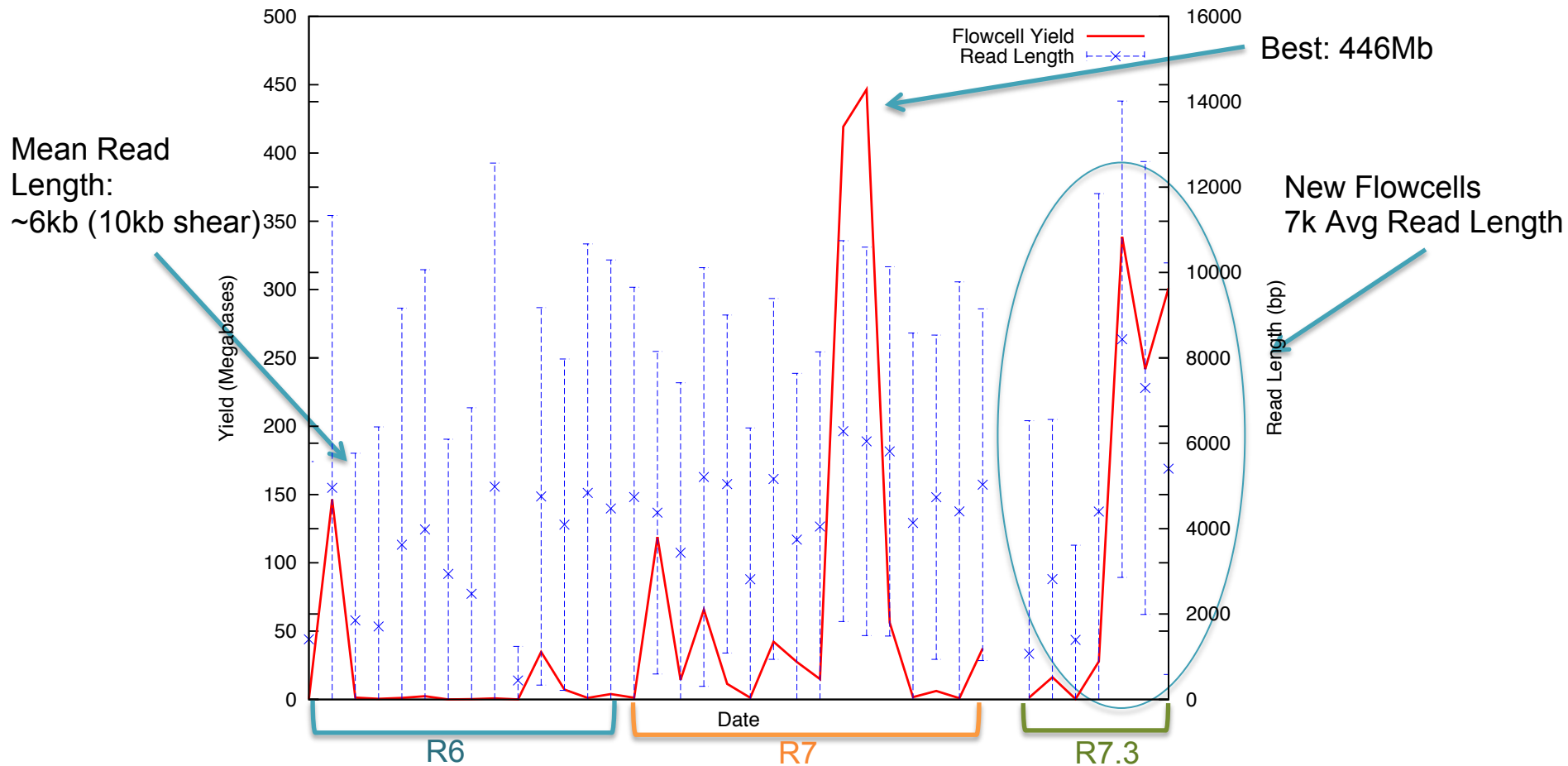
Nanopore Basecalling



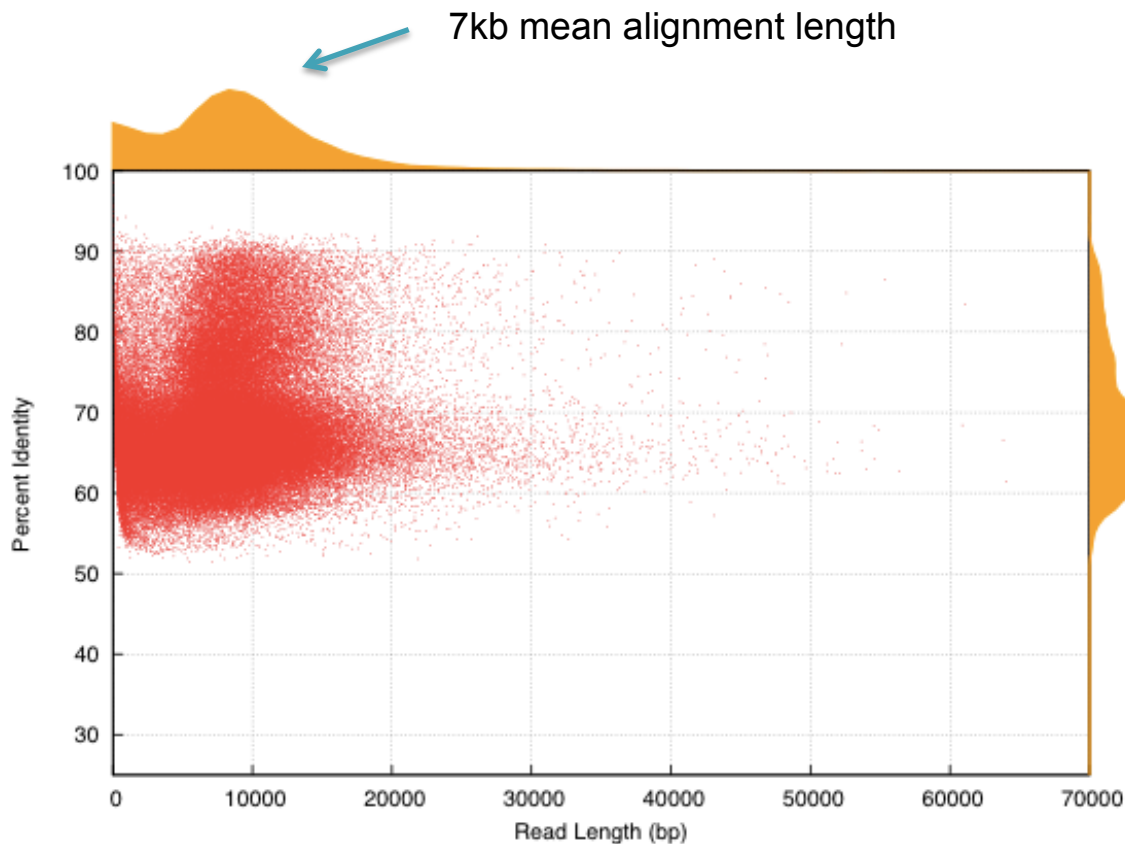
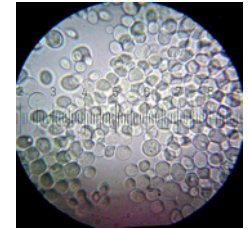
Basecalling currently performed at Amazon with frequent updates to algorithm

Our Data - Yeast W303

Oxford Flowcell Yields



Nanopore Alignments



Alignment Statistics (BLASTN)

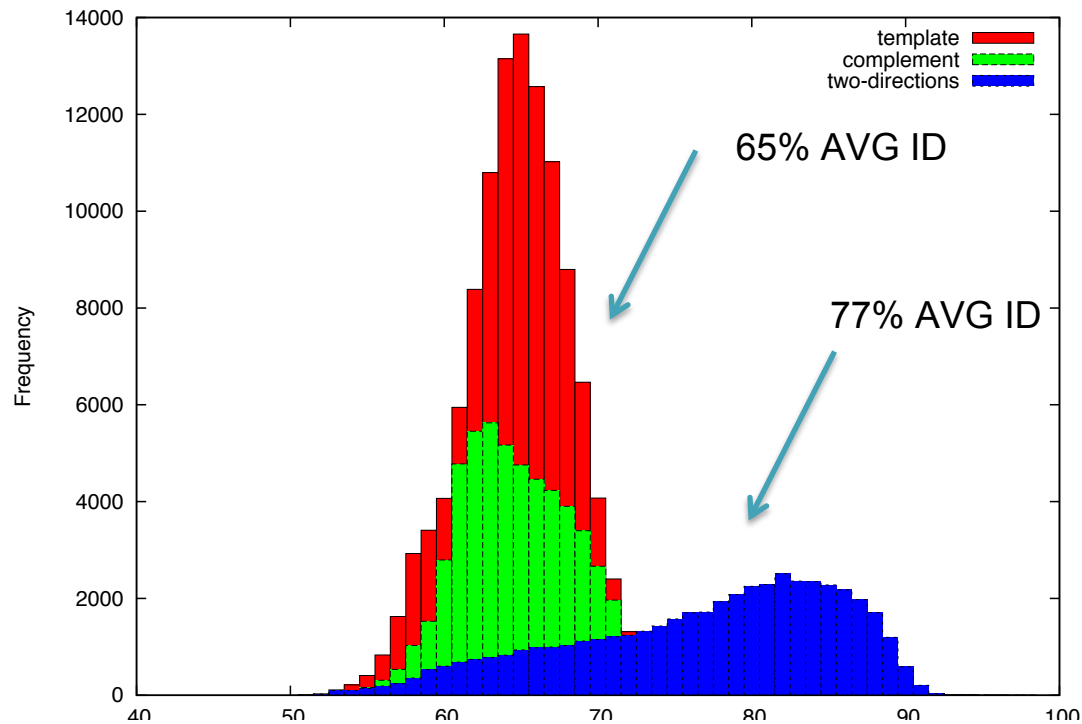
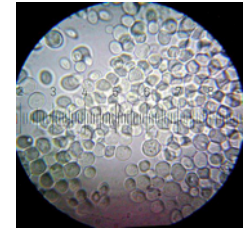
Mean alignment length at ~7kbp
Shearing targeted 10kbp
255k reads align (64%)
174x coverage

Nanopore Accuracy

Alignment Quality (BLASTN)

Of reads that align, average ~65% identity

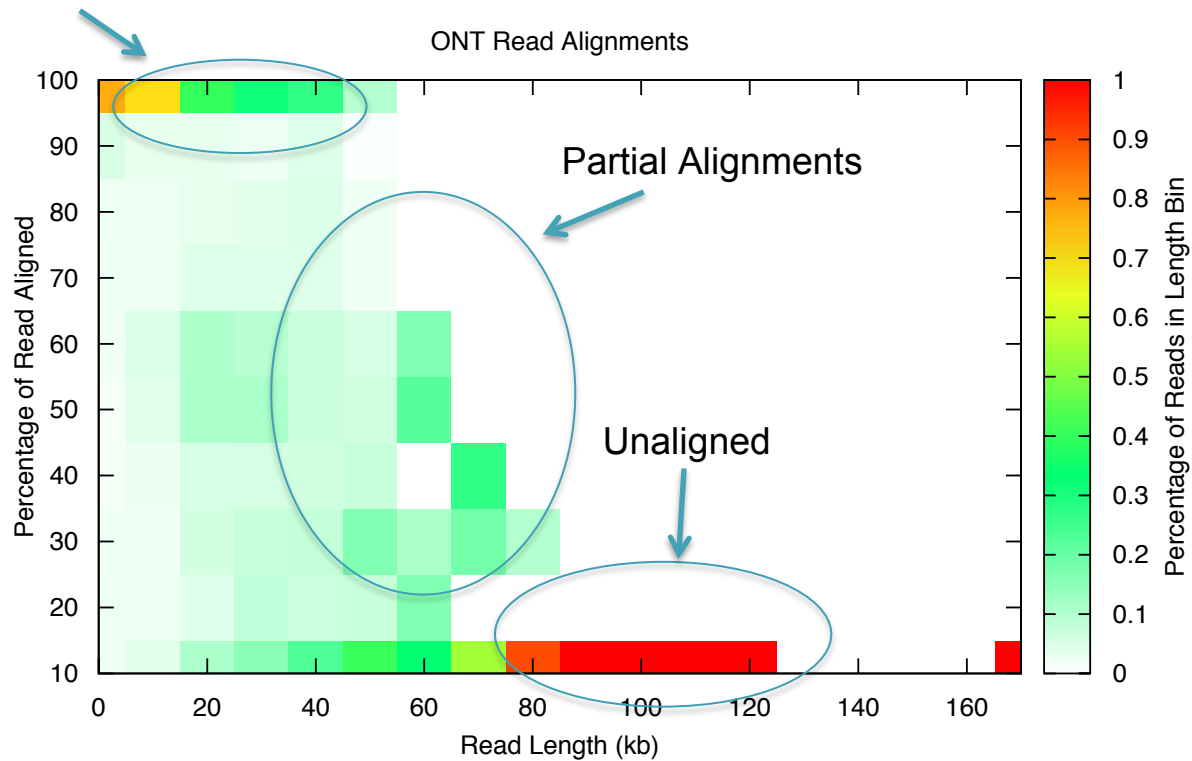
“2D base-calling” improves to ~77% identity



Nanopore Alignment Summary

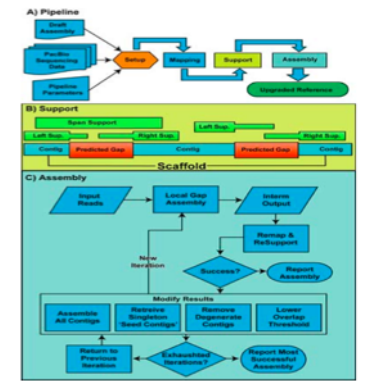
64% of the data map using BLASTN

Full Length Alignments



Long Read Correction Algorithms

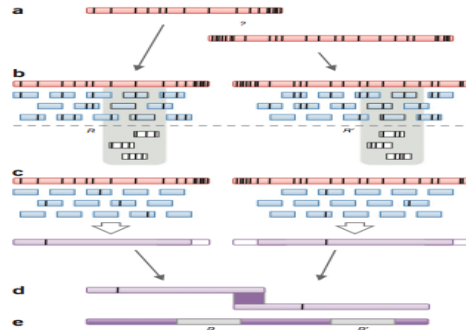
PBJelly



**Gap Filling
and Assembly Upgrade**

English *et al* (2012)
PLOS One. 7(11): e47768

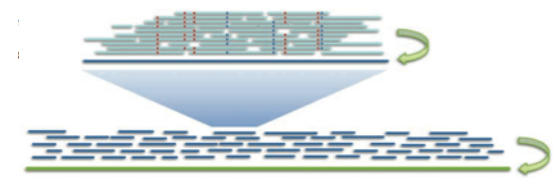
PacBioToCA & ECTools



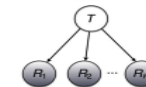
Hybrid Error Correction

Koren, Schatz, *et al* (2012)
Nature Biotechnology. 30:693–700

HGAP & Quiver



$$\Pr(\mathbf{R} | T) = \prod_k \Pr(R_k | T)$$



Quiver Performance Results Comparison to Reference Genome (<i>M. ruber</i> ; 3.1 MB; SMRT® Cells)		
	Initial Assembly	Quiver Consensus
QV	43.4	54.5
Accuracy	99.99540%	99.99964%
Differences	141	11

**LR-only Correction &
Polishing**

Chin *et al* (2013)
Nature Methods. 10:563–569

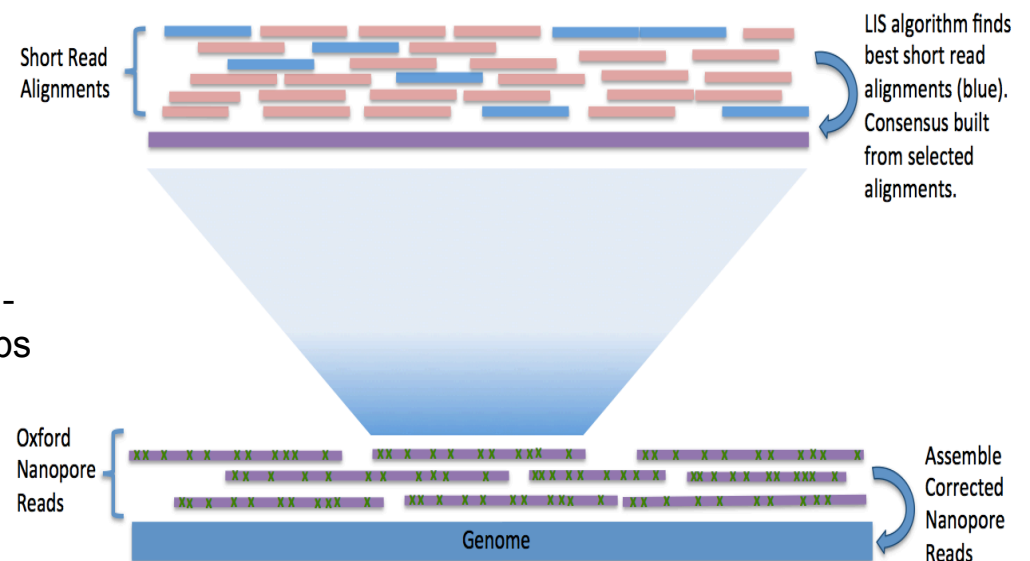
< 5x

Long Read Coverage

> 50x

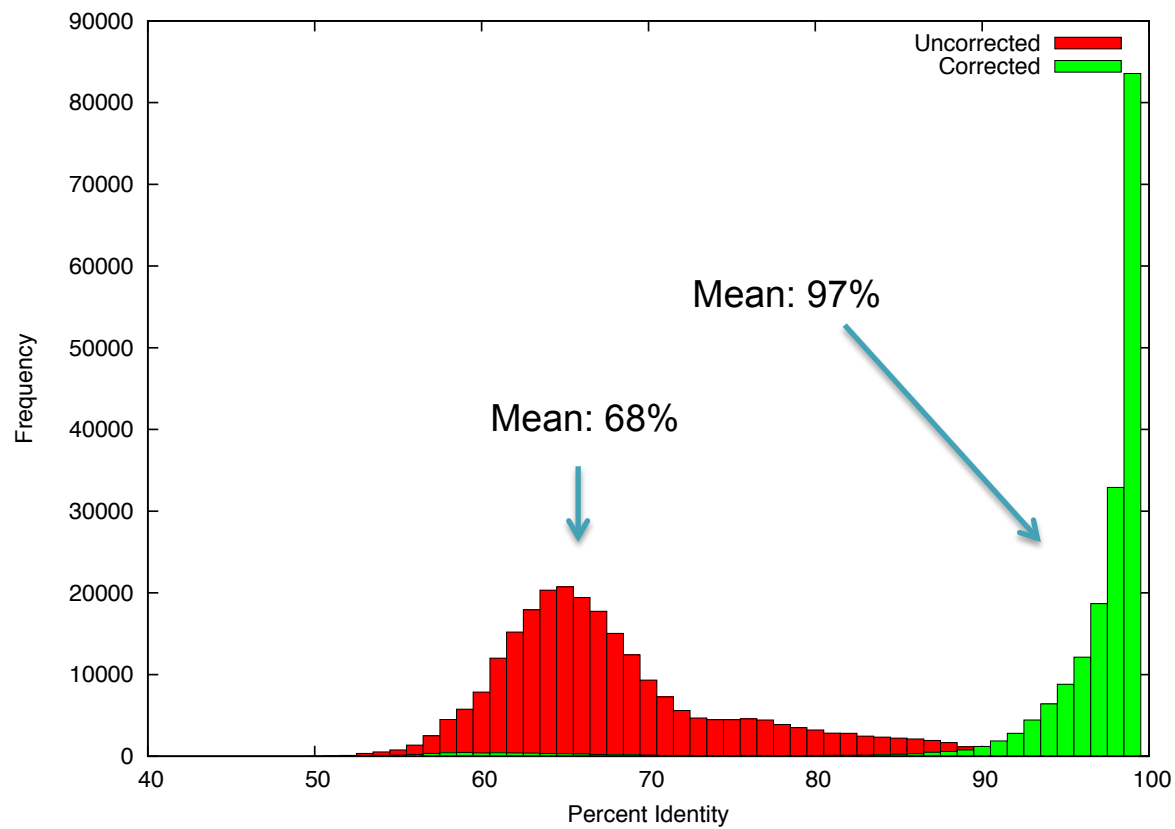
NanoCorr: Nanopore-Illumina Hybrid Error Correction

1. BLAST Miseq reads to all raw Oxford Nanopore reads
2. Select non-repetitive alignments
 - First pass scans to remove “contained” alignments
 - Second pass uses Dynamic Programming (LIS) to select set of high-identity alignments with minimal overlaps
3. Compute consensus of each Oxford Nanopore read
 - Currently using Pacbio’s pbdagcon

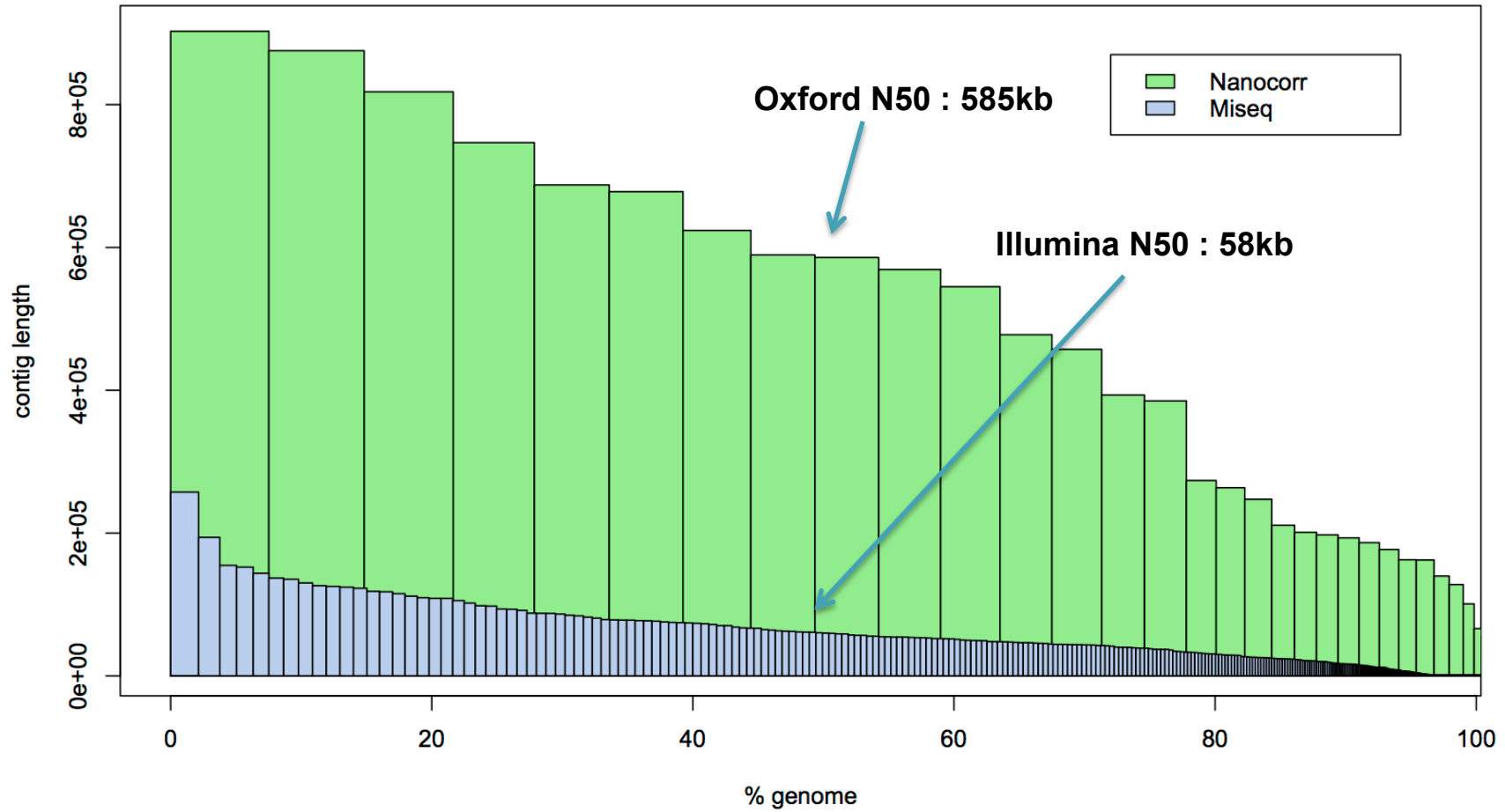


<https://github.com/jgurtowski/nanocorr>

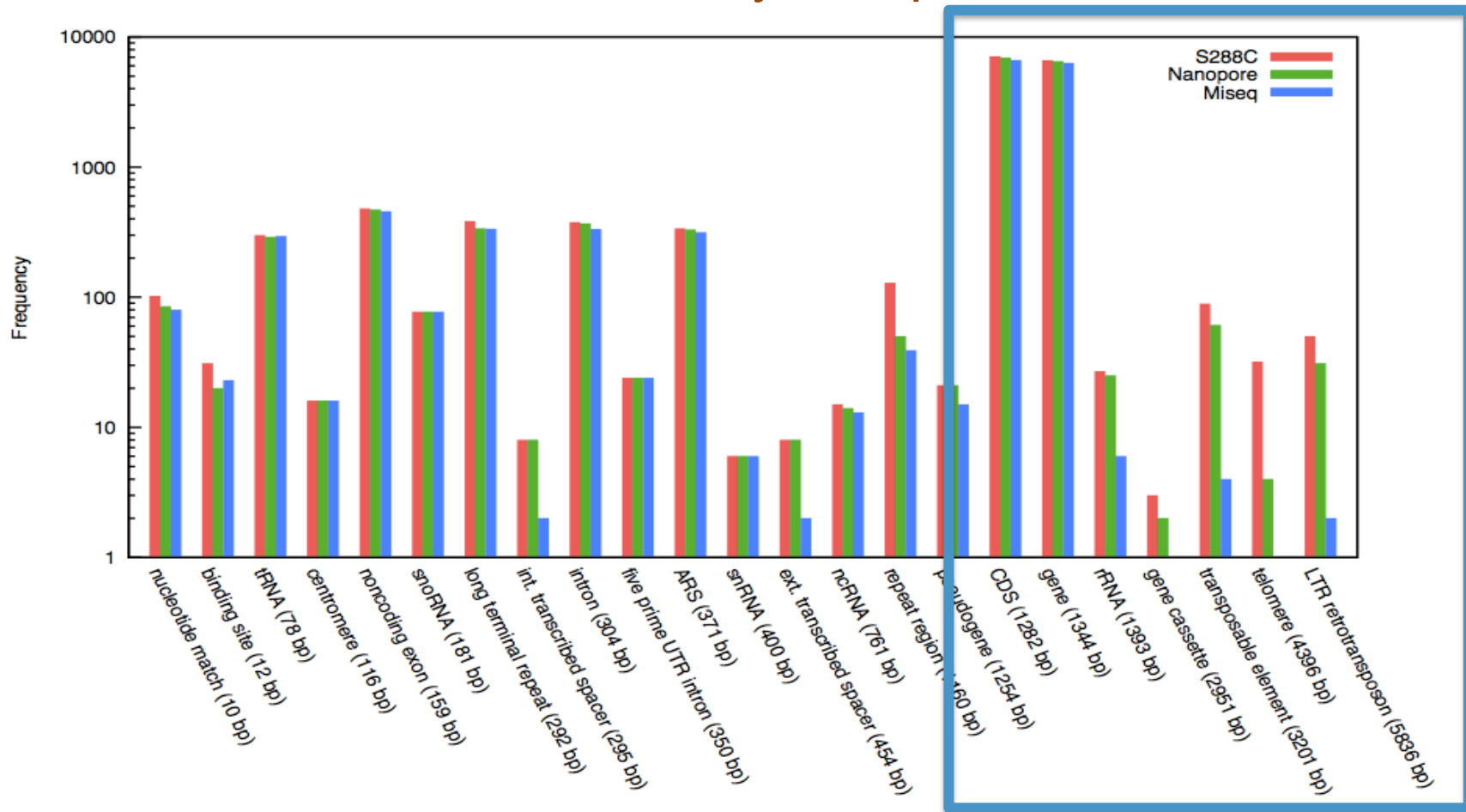
Post Correction Identity



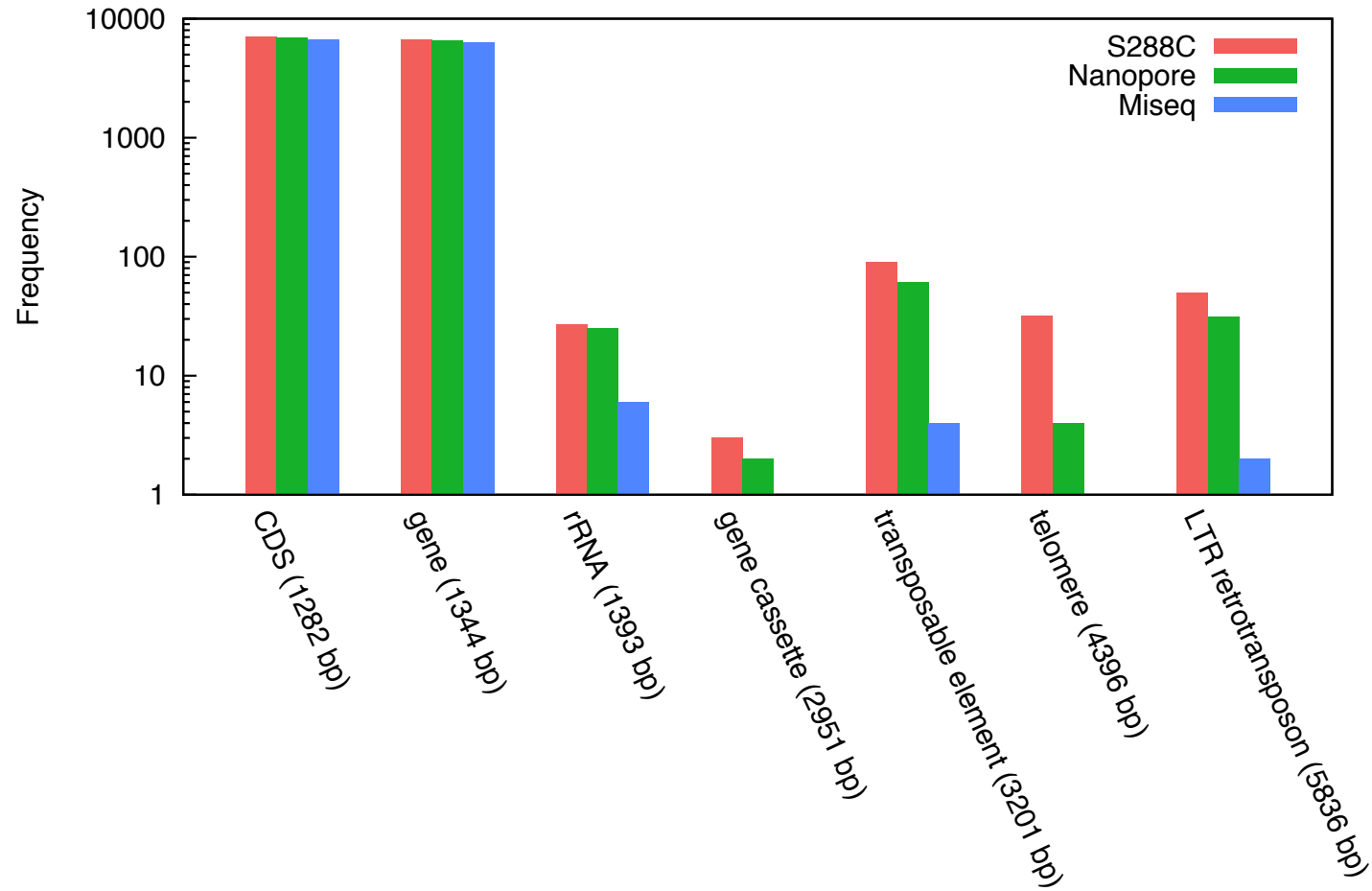
ONT vs Illumina Assembly



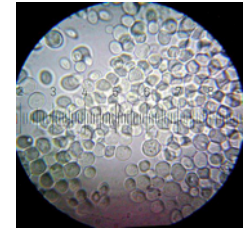
ONT Assembly Completeness



ONT Assembly Completeness



Long Read Assembly



S288C Reference sequence

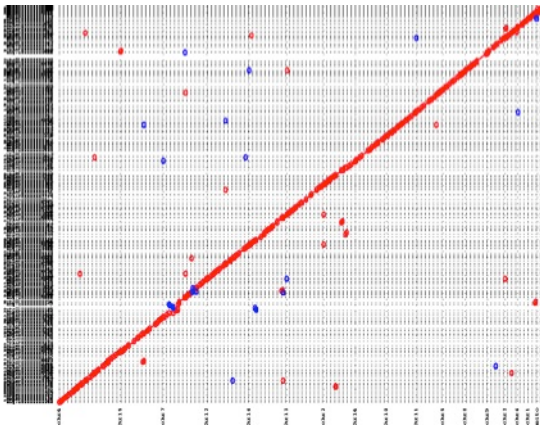
- 12.1Mbp; 16 chromo + mitochondria; N50: 924kbp

Illumina MiSeq

30x, 300bp PE (Flashed)
Celera Assembler



- 6953 non-redundant contigs
- N50: 59kb >99.9% id

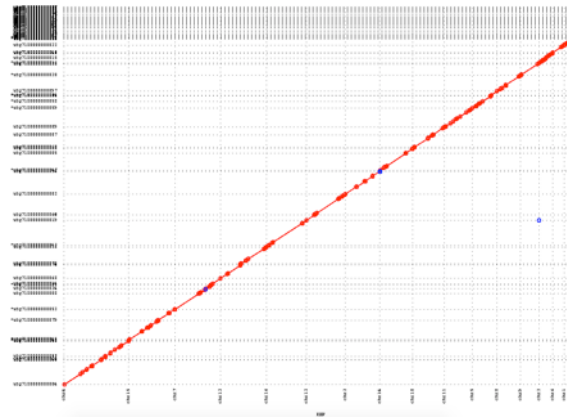


Oxford Nanopore

28x corrected reads > 7kb
NanoCorr + Celera Assembler



- 95 non-redundant contigs
- N50: 585kbp >99.78% id

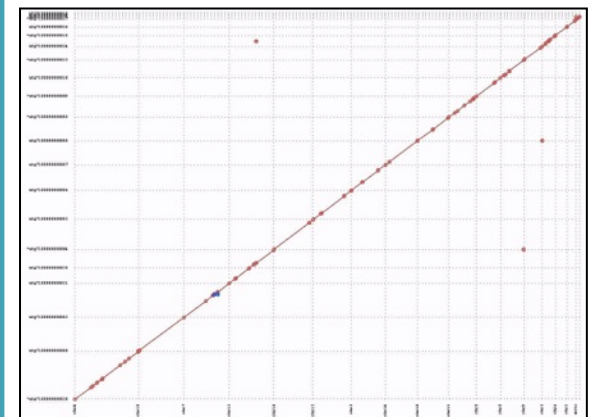


Pacific Biosciences

25x corrected reads > 10kb
HGAP + Celera Assembler



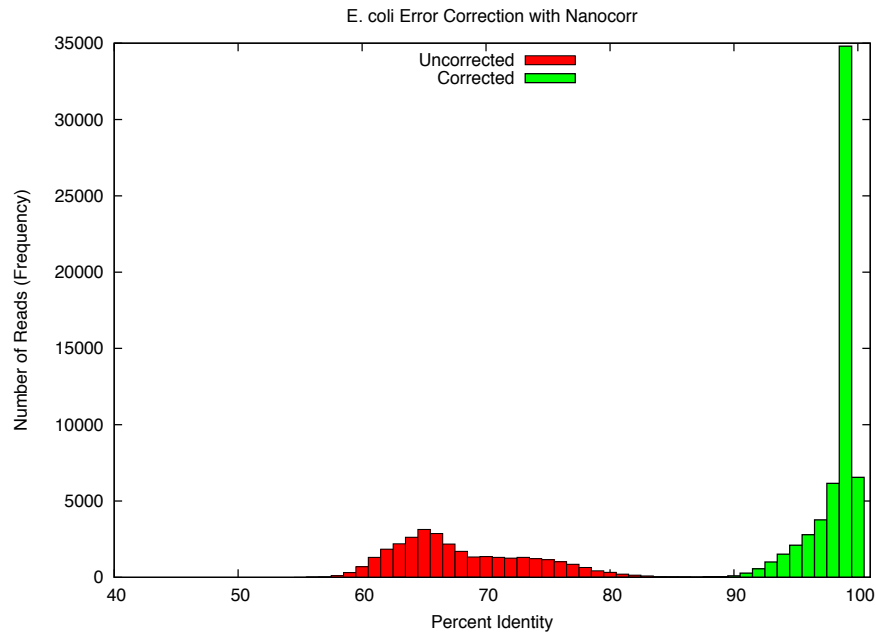
- 21 non-redundant contigs
- N50: 811kb >99.8% id



E. Coli K12 Single Contig Assembly with MinION

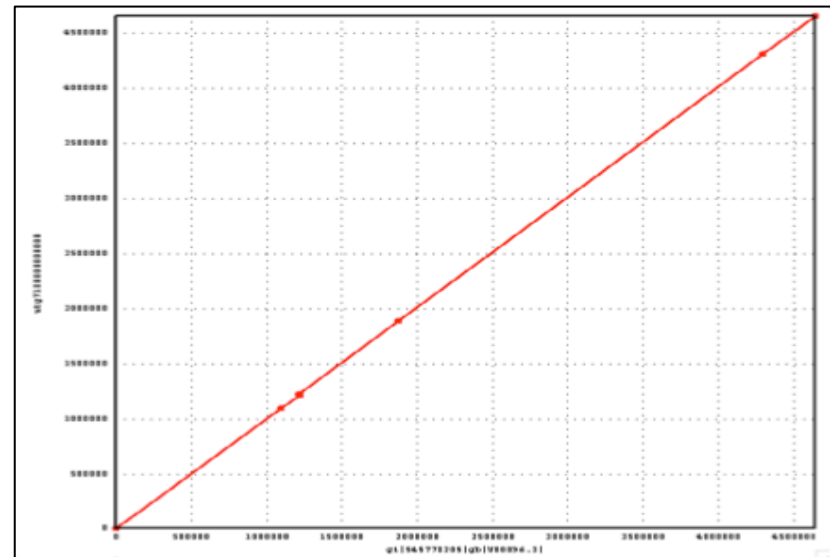
Nanocor Correction Results

145x Oxford Nanopore X 35x MiSeq



Single Contig Assembly

99.99% Identity (Pilon polishing)



Sequencing Data From:

A reference bacterial genome dataset generated on the MinION™ portable single-molecule nanopore sequencer
Joshua Quick, Aaron R Quinlan and Nicholas J Loman

Future of Oxford Nanopore



Zamin Iqbal and 5 others retweeted



GenomeWeb InSequence @InSequence · Oct 20

Oxford Nanopore shows off **PromethION** at **ASHG**. #ASHG14 #nanopore



Acknowledgements



Michael Schatz

Dick McCombie

Sara Goodwin

Schatz Lab



Oxford Nanopore Sequencing and de novo Assembly of a Eukaryotic Genome

Sara Goodwin , James Gurtowski , Scott Ethe-Sayers , Panchajanya Deshpande ,
Michael Schatz , W Richard McCombie

doi: <http://dx.doi.org/10.1101/013490>