Real Time minION data analysis: Run and Read Until

Matt Loose

#1

The future is here?


“Wondering how you can fit MinION into your existing workflows is like sitting in front a space ship and wondering how you’re going to use it to commute to work; it’s like sitting in front of a time machine, and wondering if you could get to the shops before they close.”
@BioMickWatson

5-10/20 Gb

50 - 100 Gb

3-6 Tb

#2

Run Monitoring

minoTour (v0.7)

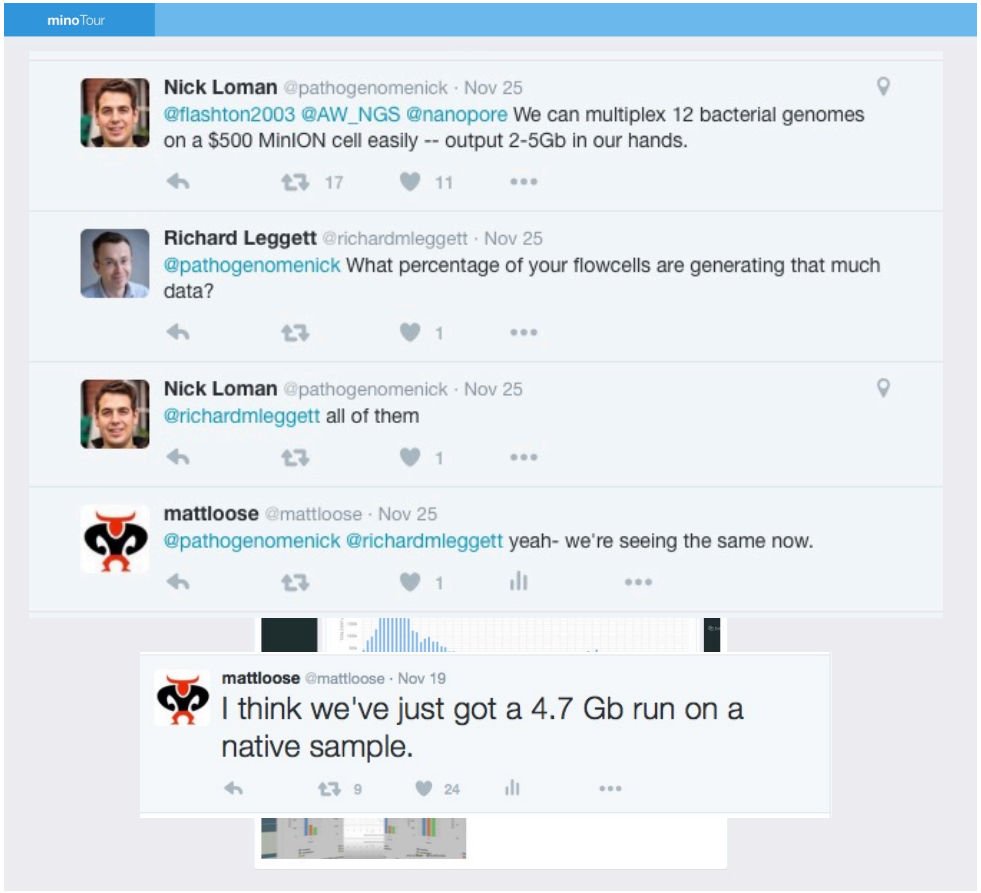
What is it?

Launched September 2014 (30 b/s)

Slowed down March 2016 (250 b/s)

Struggled with new MinKNOW

Now available at GitHUB for local install....



inoTour(v0.7)

Real Time Analysis

Processes:

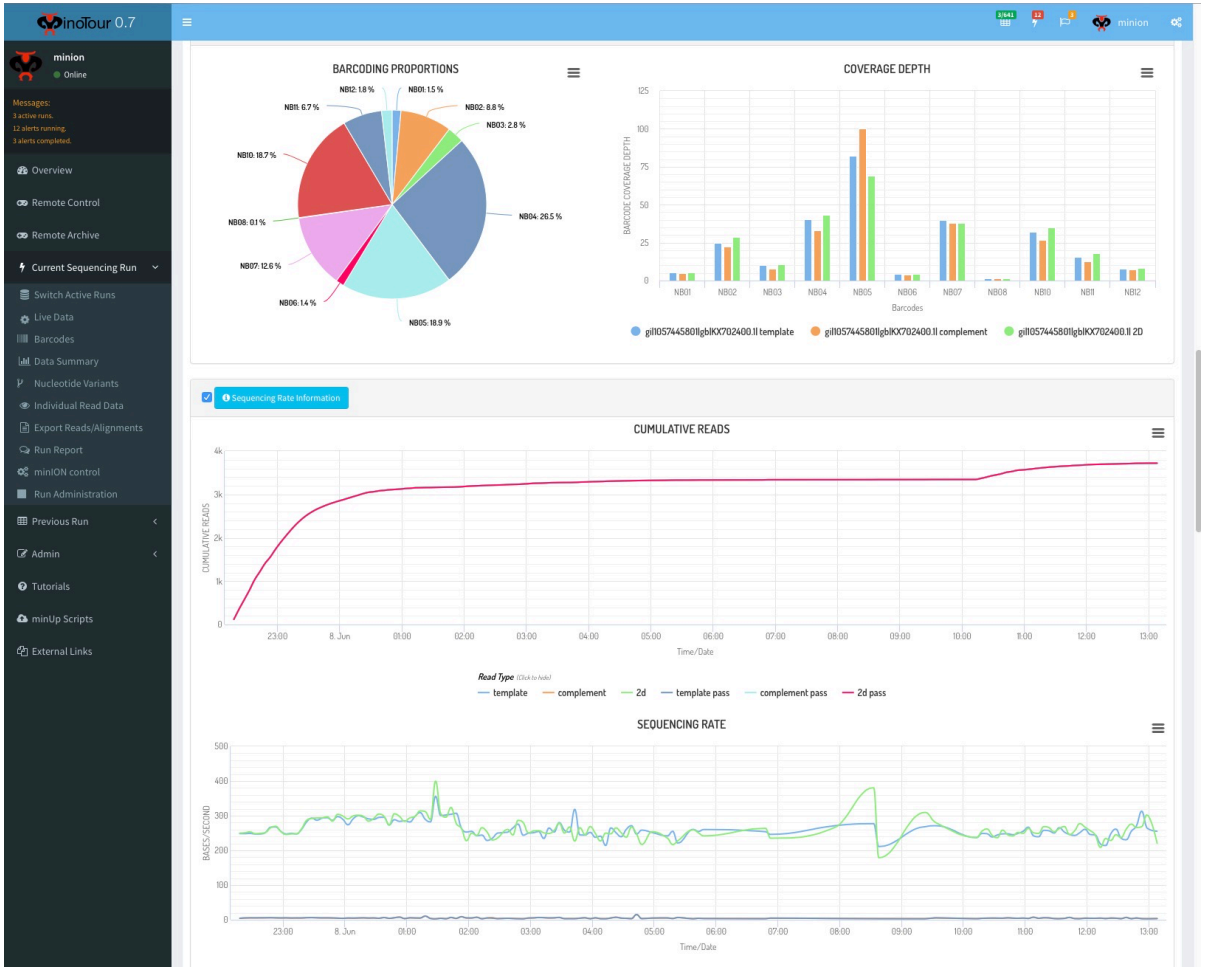
- minKNOW
- nanonet
- metrichor

Handles:

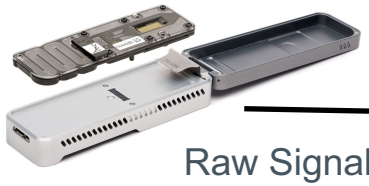
- Barcodes
- Alignment to Reference

Provides:

- Metrics



Monitoring Runs



Raw Signal

minKNOW

2d or 1d without base calling

Integrated Basecalling

1d Only Raw Signal FastQ

Raw Signal Events (opt)

metrichor

Nanonet

Third Party Base Callers

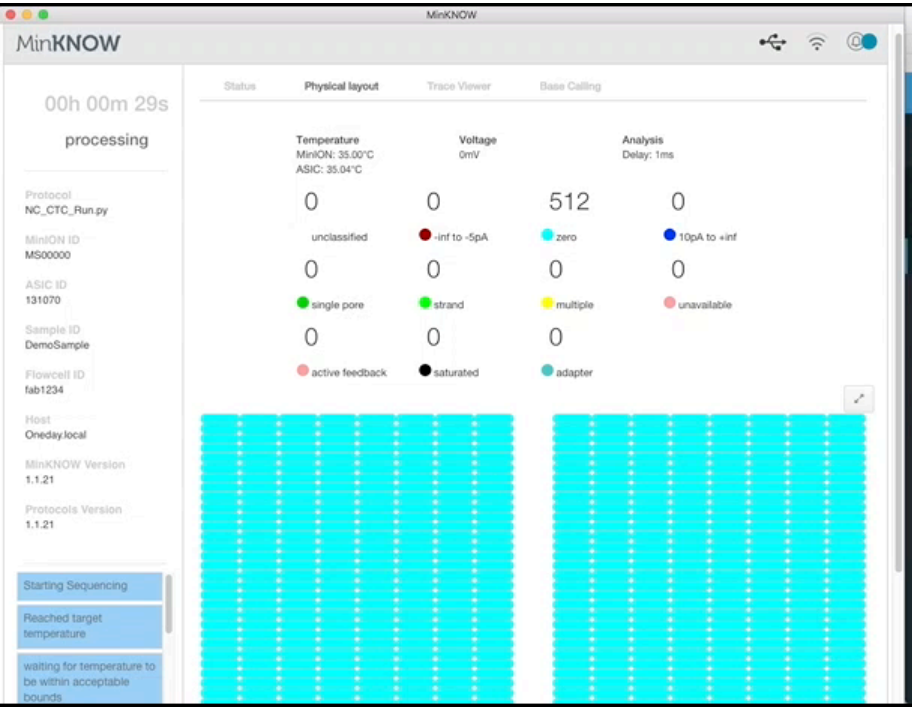
Fasta/Q

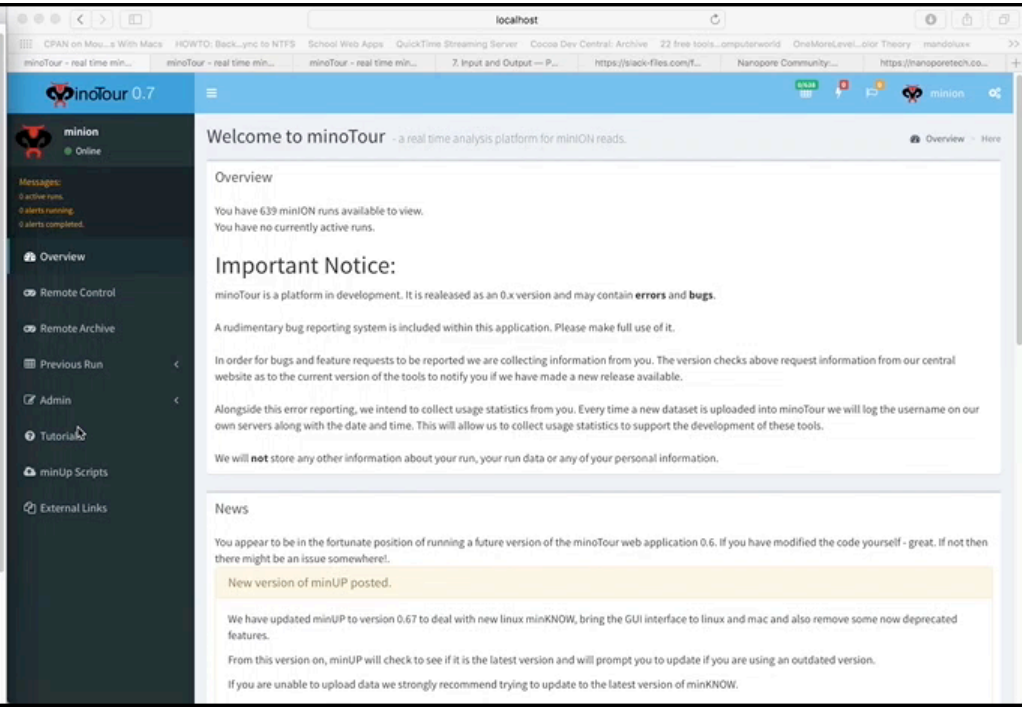
1d/2d Raw Signal Events FastQ

1d/2d Raw Signal Events (opt) FastQ

inoTour

minoTour keeps up!



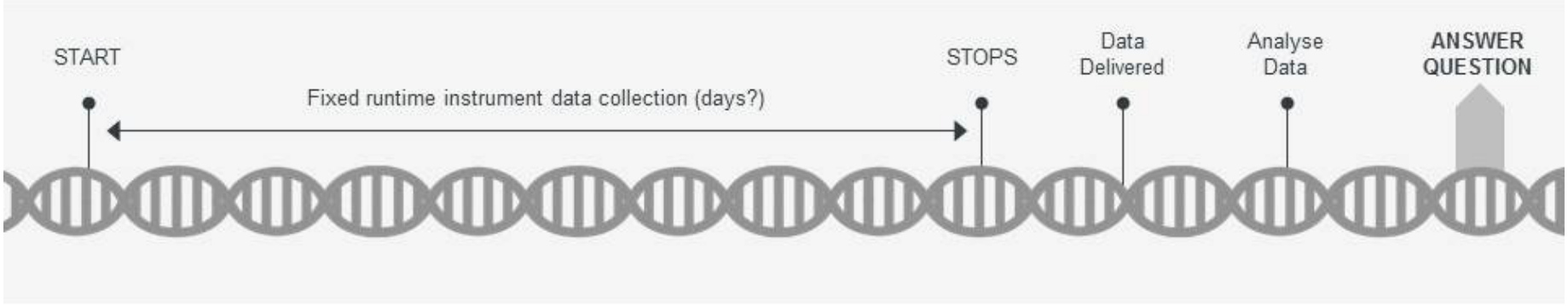


#3

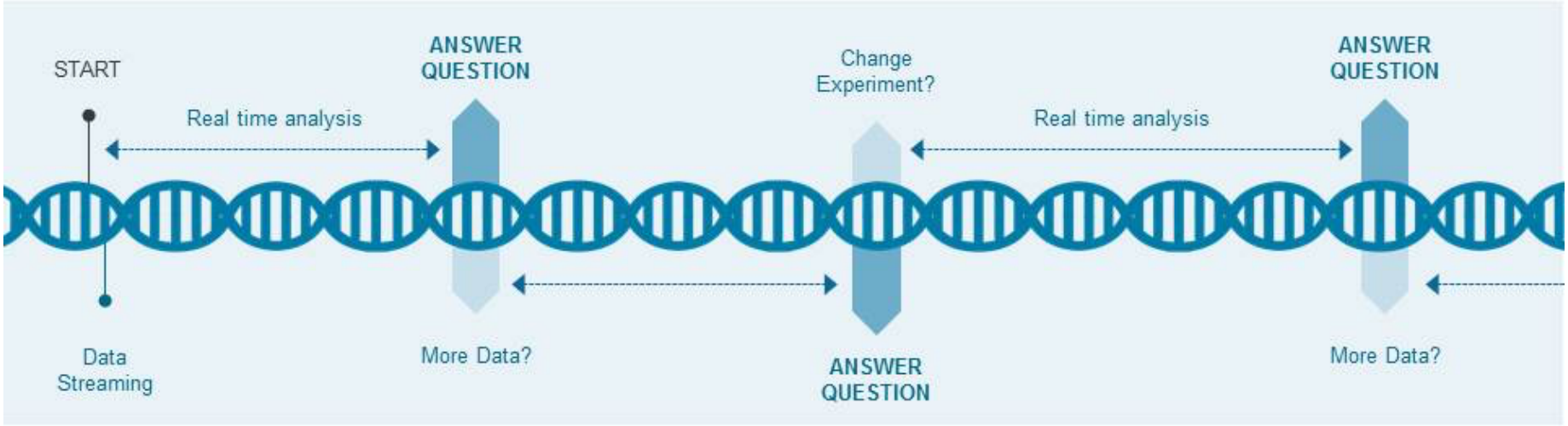
Run Until

Run Until…

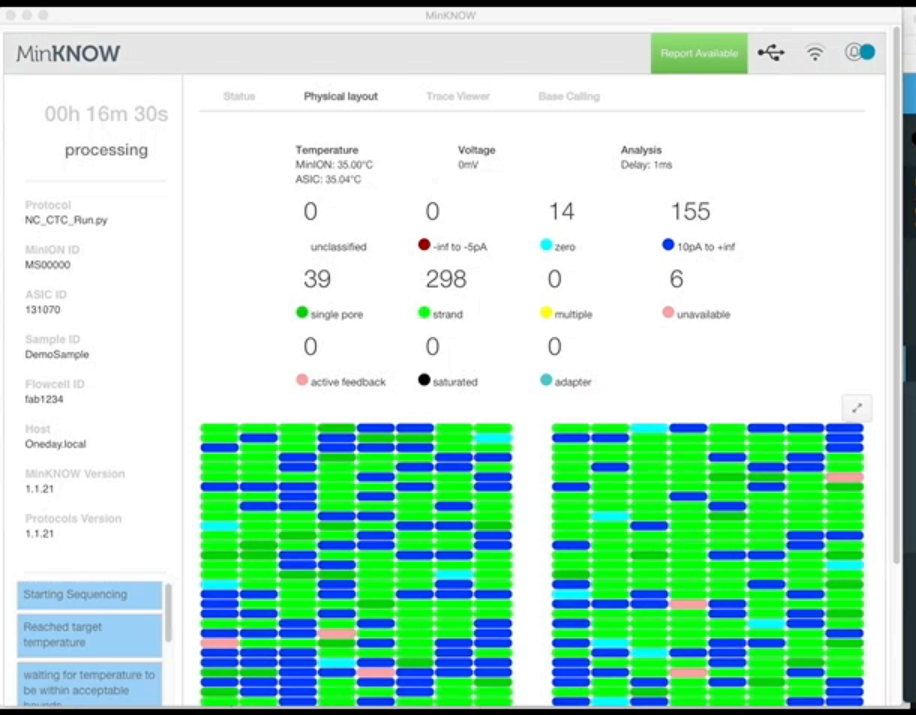
Traditional Workflow

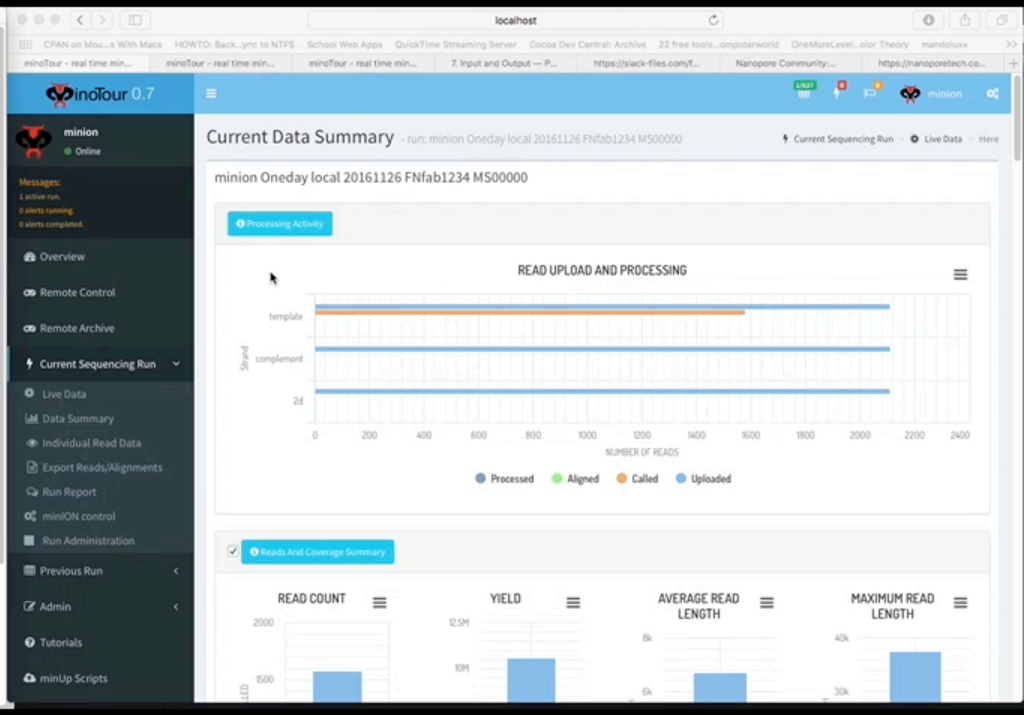


Run Until…



© Copyright 2015 Oxford Nanopore Technologies


MinKNOW
MinKNOW
Report Available
00h 16m 30s
processing
Status
Physical layout
Trace Viewer
Base Calling
Temperature
MinION: 35.00°C
ASIC: 35.04°C
Voltage
0mV
Analysis
Delay: 1ms
0
unclassified
0
-inf to -5pA
14
zero
155
10pA to +inf
39
single pore
298
strand
0
multiple
6
unavailable
0
active feedback
0
saturated
0
adapter
Protocol
NC_CTC_Run.py
MinION ID
MS00000
ASIC ID
131070
Sample ID
DemoSample
Flowcell ID
fab1234
Host
Oneday.local
MinKNOW Version
1.1.21
Protocols Version
1.1.21
Starting Sequencing
Reached target temperature
waiting for temperature to be within acceptable


localhost
minoTour 0.7
minion
Online
Messages:
1 active run.
0 alerts running.
0 alerts completed.
Overview
Remote Control
Remote Archive
Current Sequencing Run
Live Data
Data Summary
Individual Read Data
Export Reads/Alignments
Run Report
minION control
Run Administration
Previous Run
Admin
Tutorials
minUp Scripts
Current Data Summary
run: minion Oneday local 20161126 FNfab1234 MS00000
minion Oneday local 20161126 FNfab1234 MS00000
Processing Activity
READ UPLOAD AND PROCESSING
template
complement
2d
Strand
0
200
400
600
800
1000
1200
1400
1600
1800
2000
2200
2400
NUMBER OF READS
Processed
Aligned
Called
Uploaded
Reads And Coverage Summary
READ COUNT
YIELD
AVERAGE READ LENGTH
MAXIMUM READ LENGTH

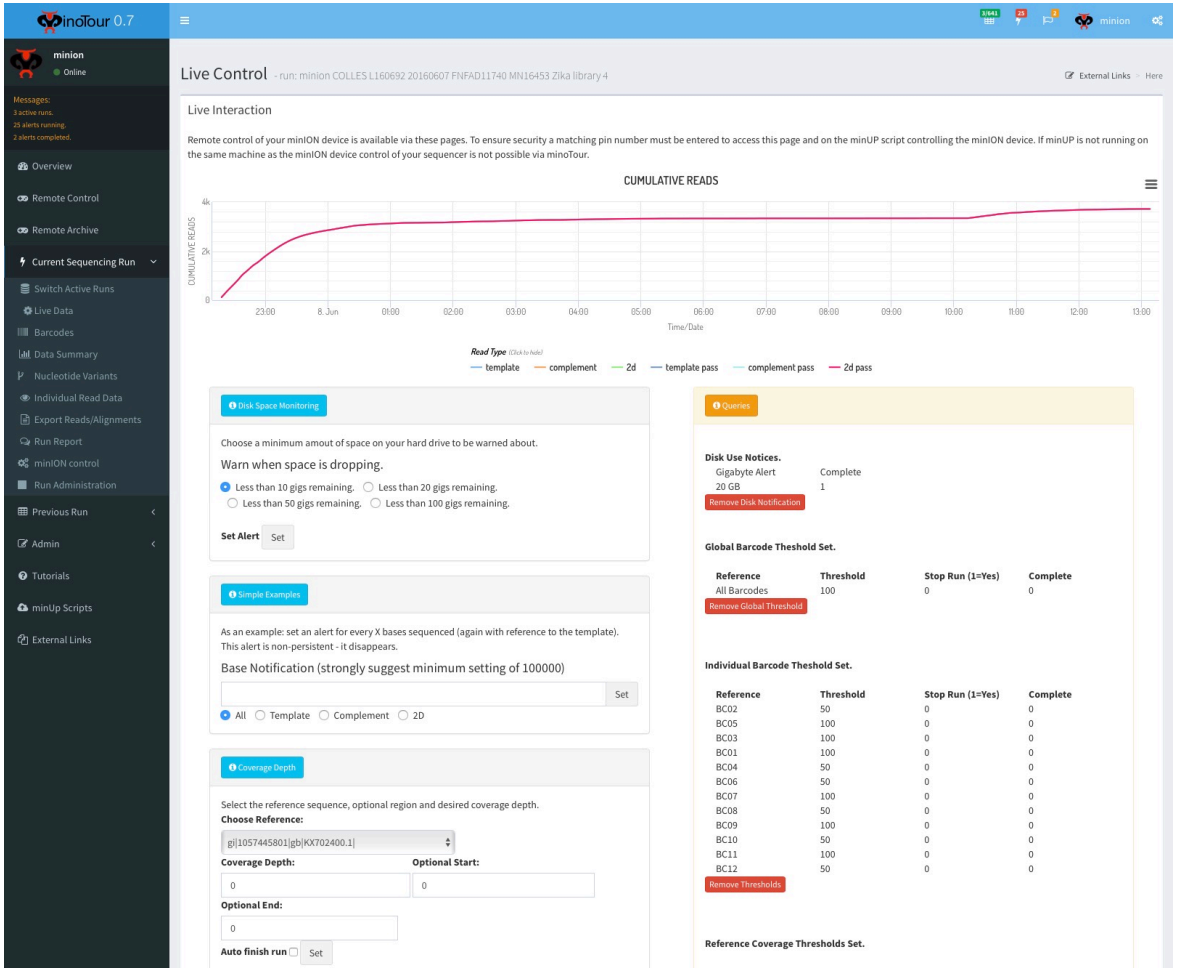
‘Run Until’

Notifications (optional stop):

- Global Coverage Depth
- Targeted Region Depth
- Depth Per Barcode (same reference only)

Notifications:

- Total number of bases sequenced
- Drive Space Warnings

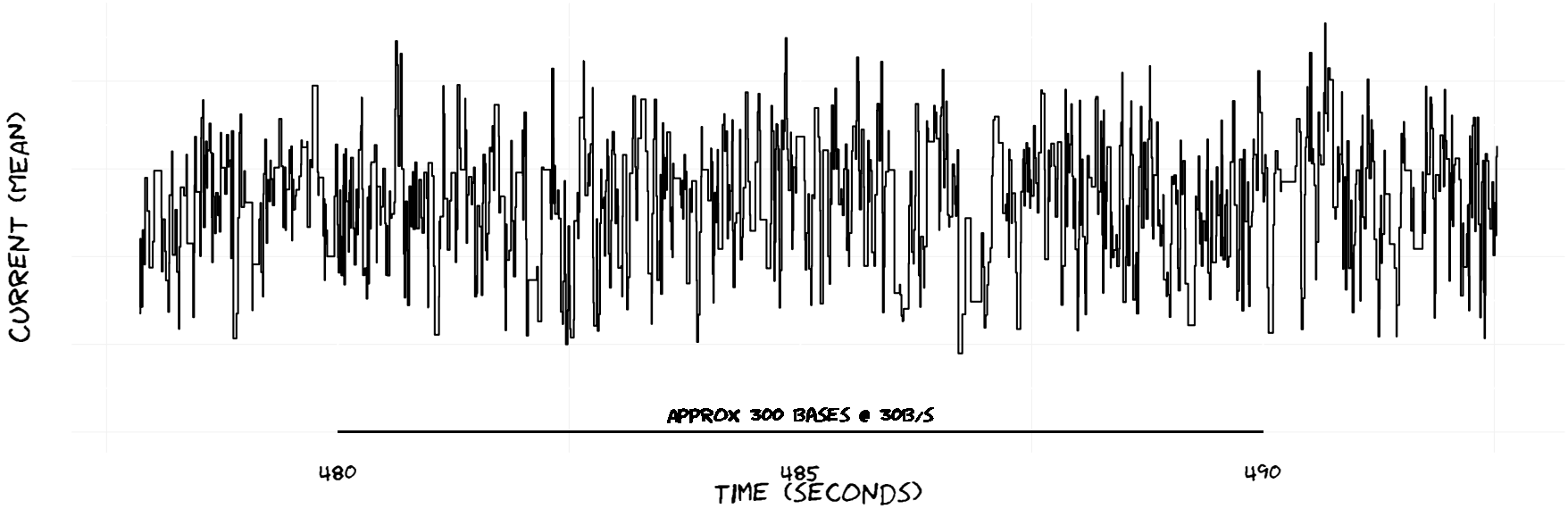


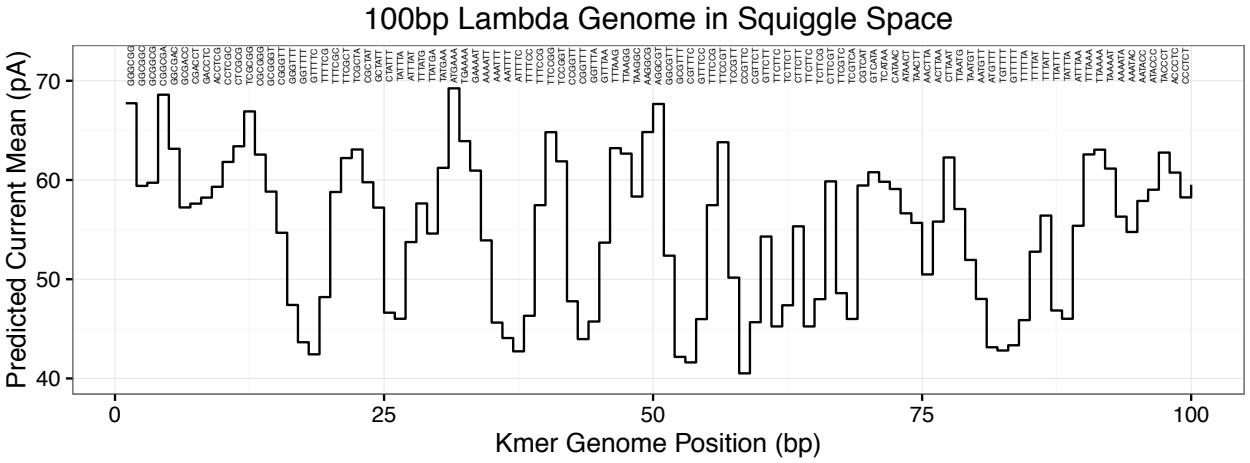
#4

Read Until

Read Until – Selective Sequencing

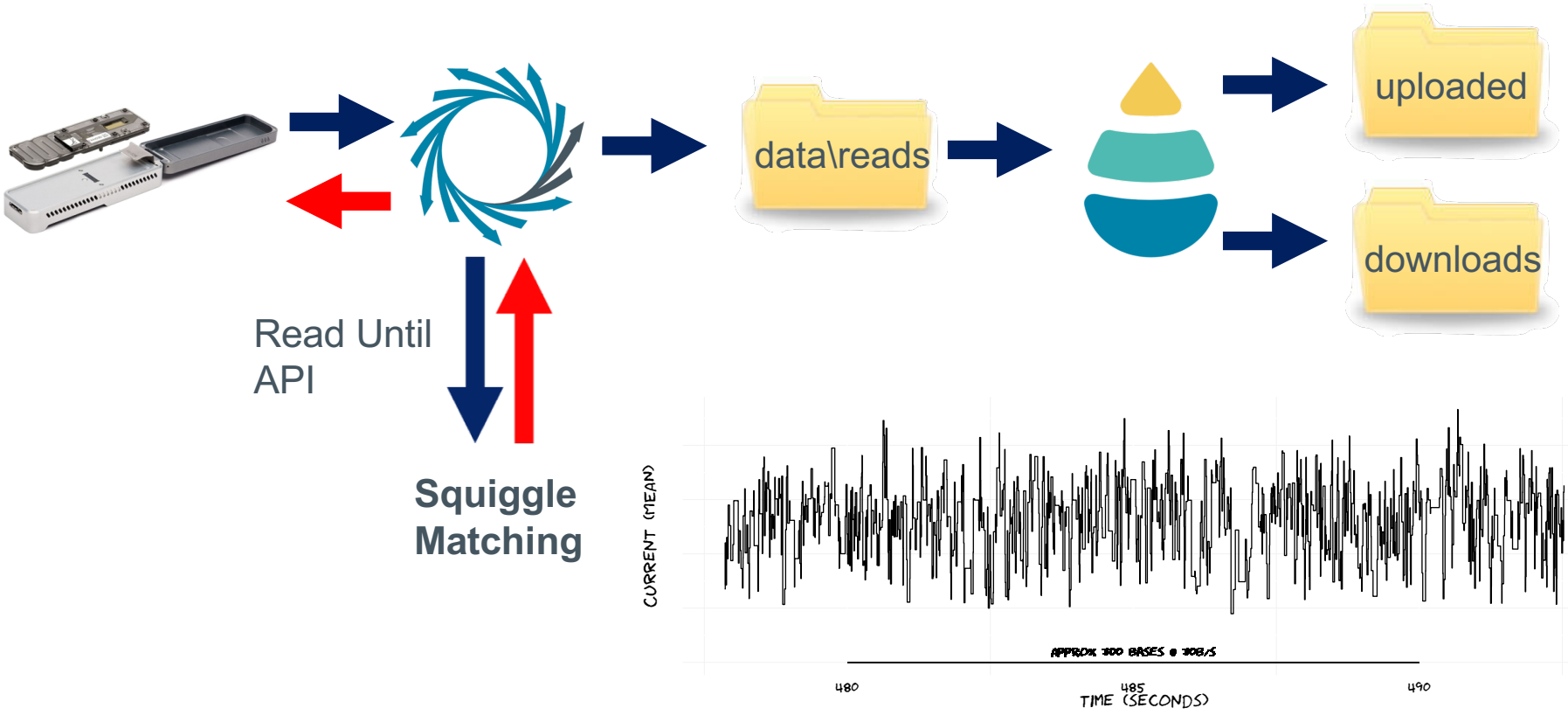
Sequencing Workflow

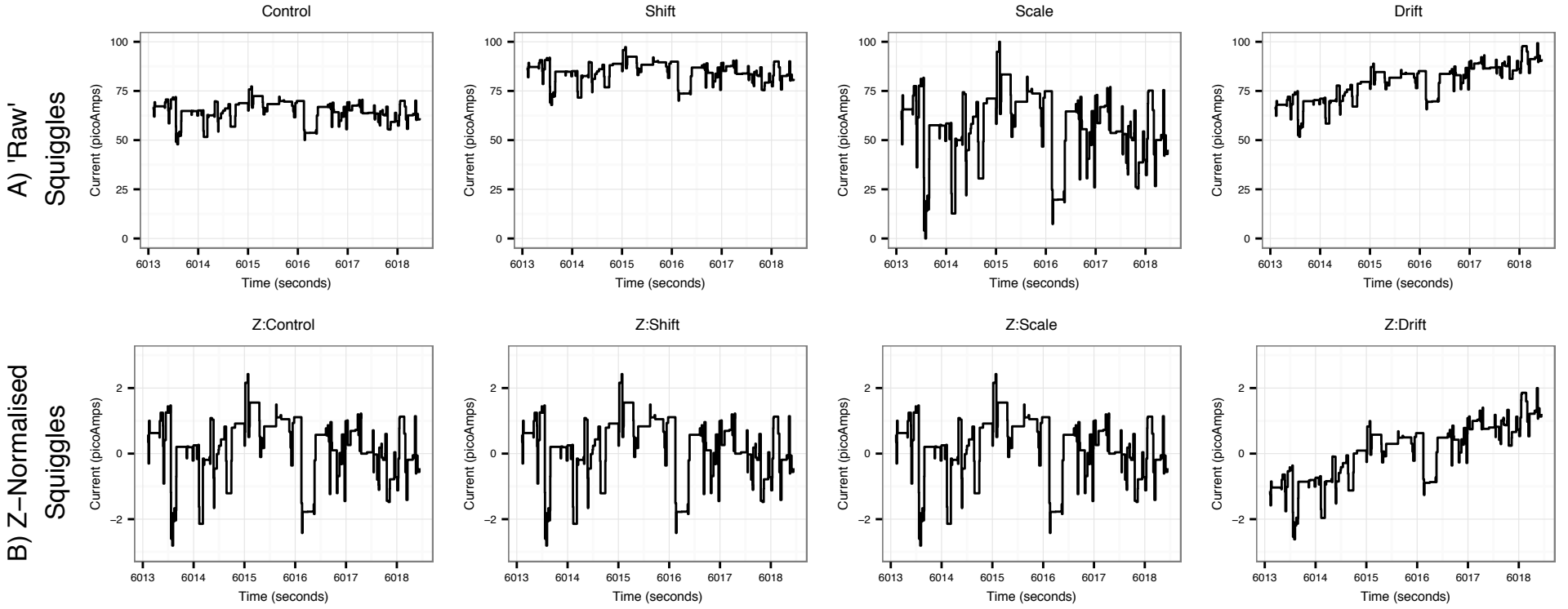


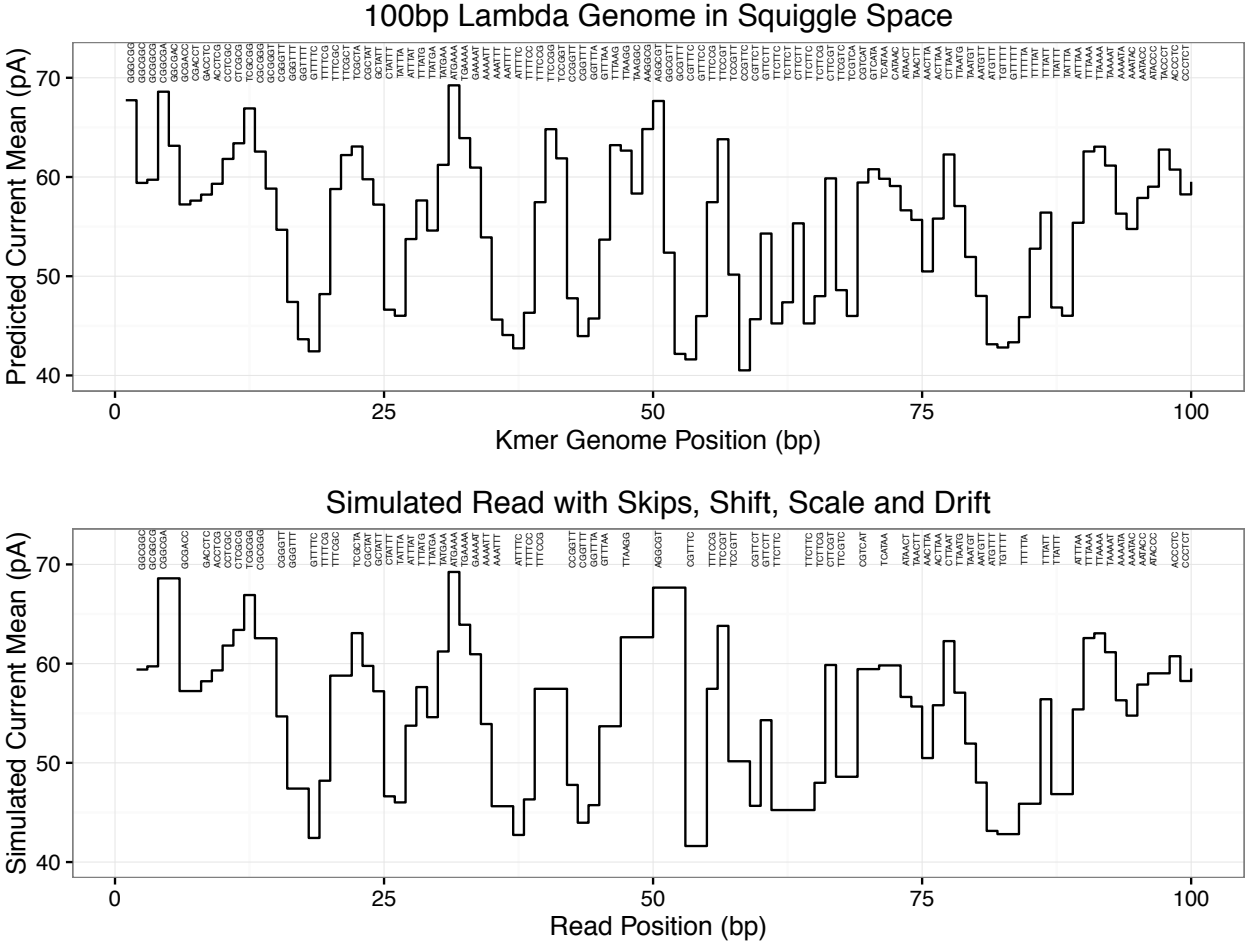


GGGCGG	67.73595
GGCGGC	59.405716
GCGGCG	59.723809
CGGCGA	68.596572
GGCGAC	63.144473
GCGACC	57.233718
CGACCT	57.61982
GACCTC	58.224973
ACCTCG	59.316305
CCTCGC	61.821412

Read Until Workflow

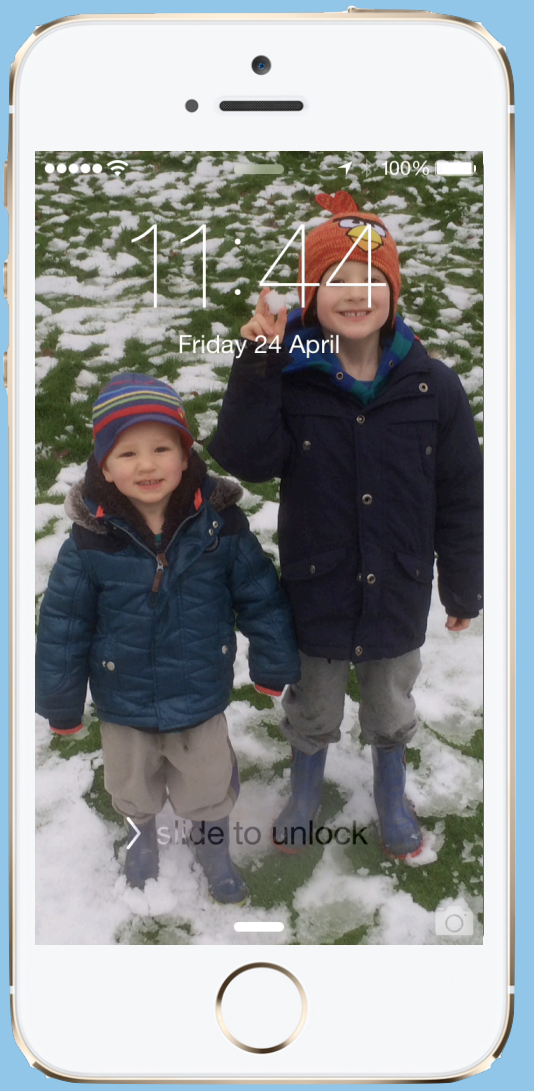



A) 'Raw' Squiggles
Control
Shift
Scale
Drift
Current (picoAmps)
Time (seconds)
B) Z–Normalised Squiggles
Z:Control
Z:Shift
Z:Scale
Z:Drift


100bp Lambda Genome in Squiggle Space
Predicted Current Mean (pA)
70
60
50
40
0
25
50
75
100
Kmer Genome Position (bp)
Simulated Read with Skips, Shift, Scale and Drift
Simulated Current Mean (pA)
70
60
50
40
0
25
50
75
100
Read Position (bp)

#5

How To Compare Squiggles?



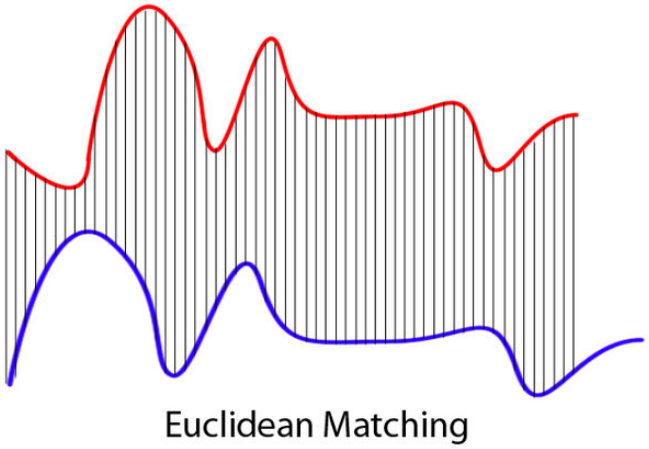
Dynamic Time Warping

Calculates the minimal distance between two paths which vary in time

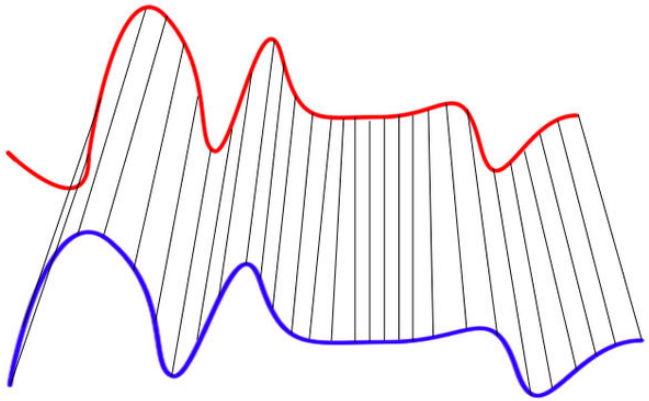
Guaranteed to find the optimal match

Computationally Expensive

Returns: Distance, Cost, Path



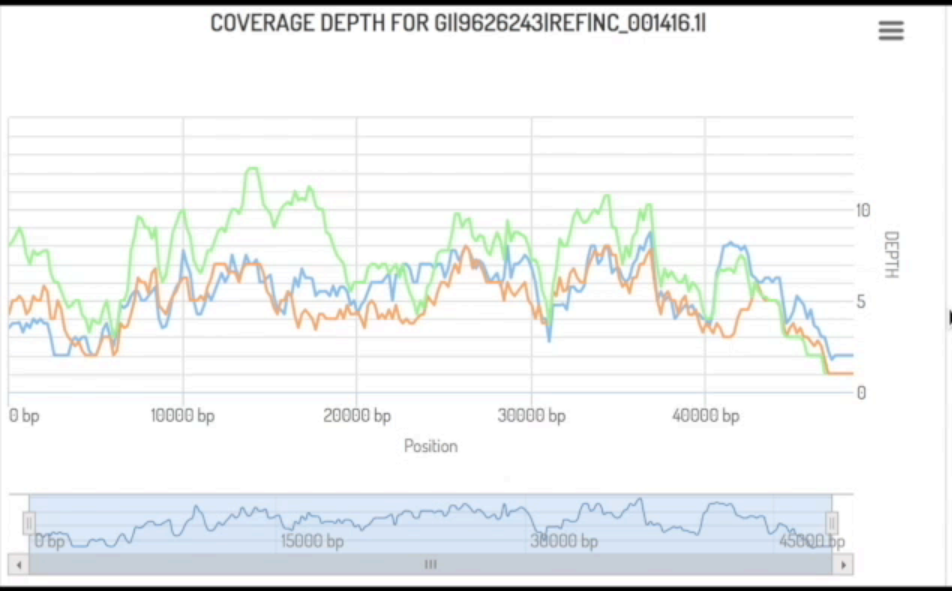
Euclidean Matching

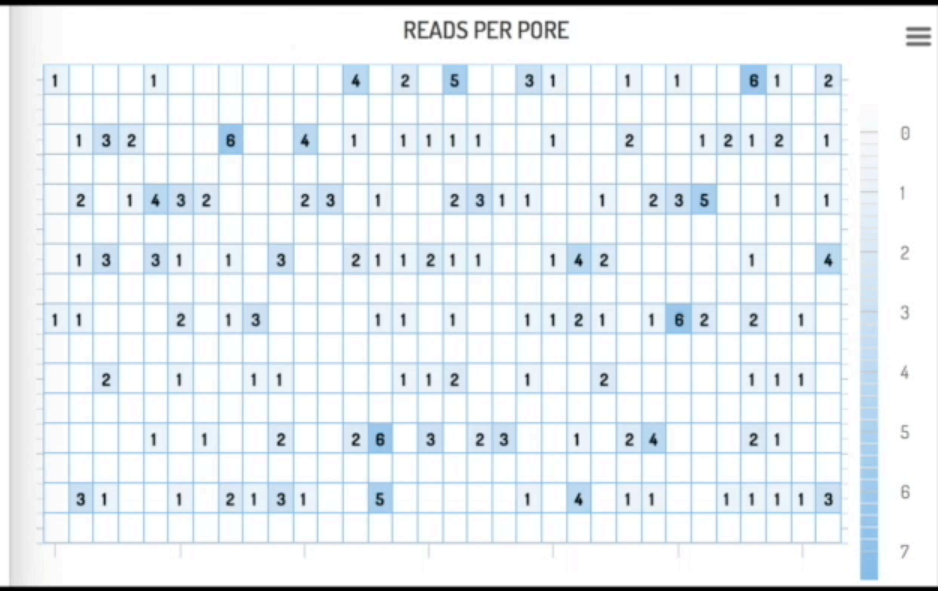


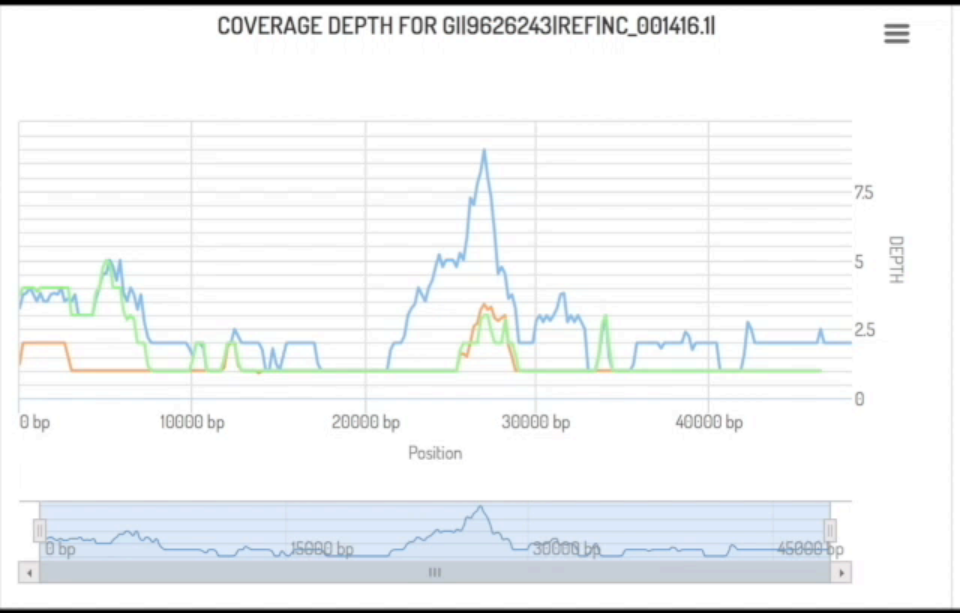
Dynamic Time Warping Matching

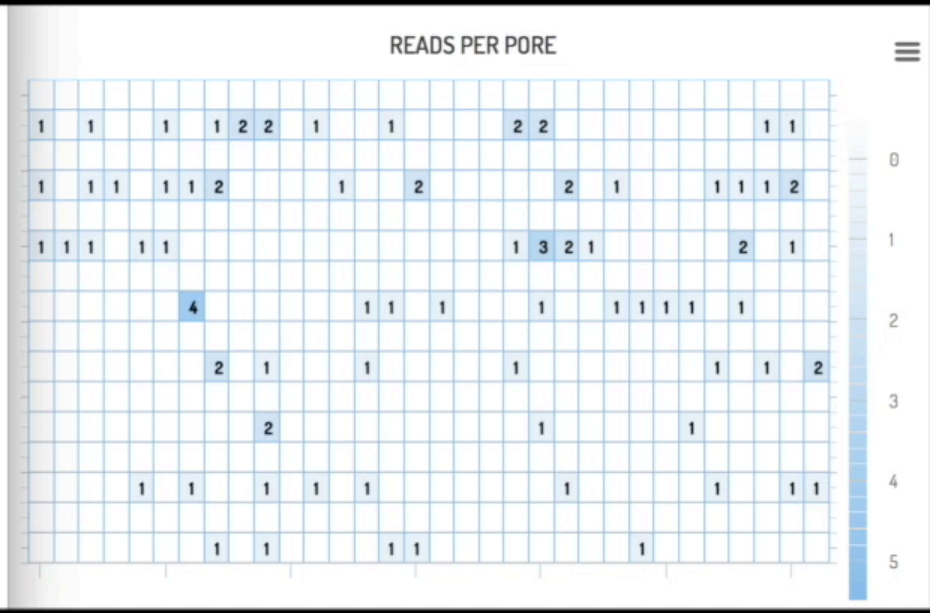
#6

Does Read Until Actually Work?

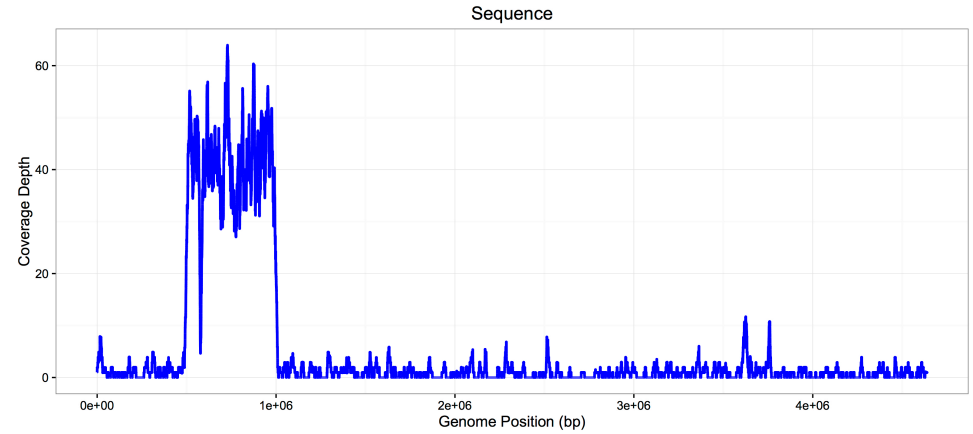

COVERAGE DEPTH FOR GI|9626243|REF|NC_001416.1|
10
5
0
DEPTH
0 bp
10000 bp
20000 bp
30000 bp
40000 bp
Position
0 bp
15000 bp
30000 bp
45000 bp

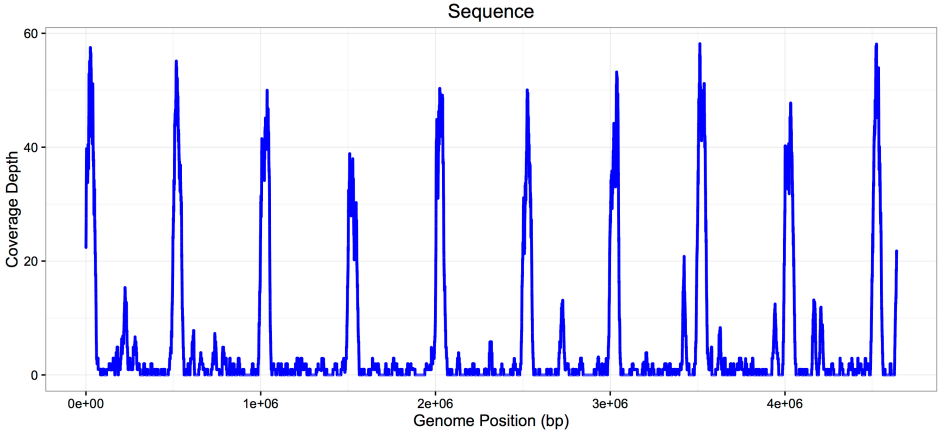

READS PER PORE
0
1
2
3
4
5
6
7

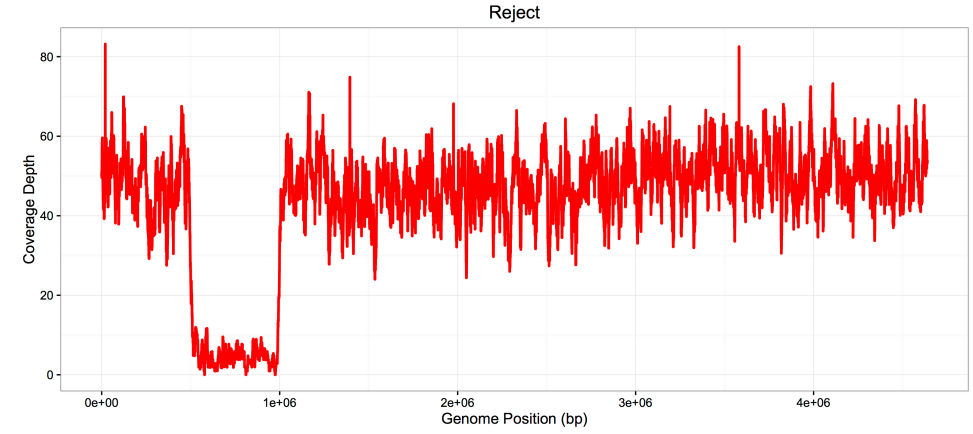

COVERAGE DEPTH FOR GI|9626243|REF|NC_001416.1|
7.5
5
2.5
0
DEPTH
0 bp
10000 bp
20000 bp
30000 bp
40000 bp
Position
0 bp
15000 bp
30000 bp
45000 bp

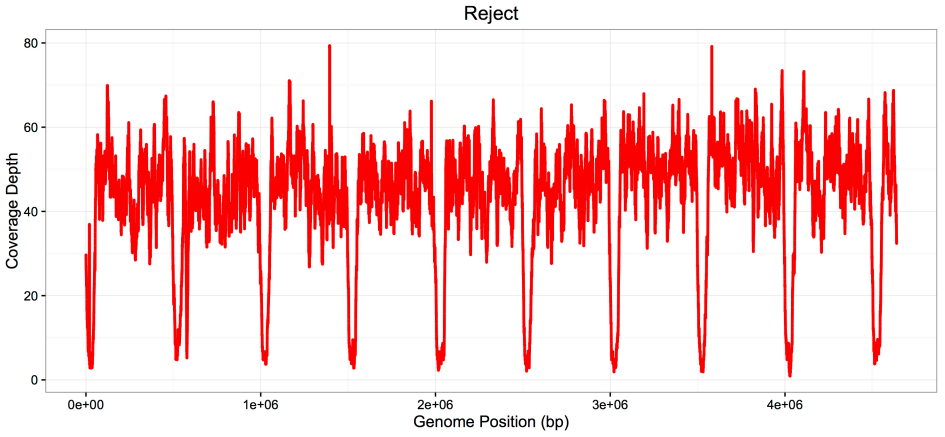

READS PER PORE
0
1
2
3
4
5

Selective Sequencing from a Background







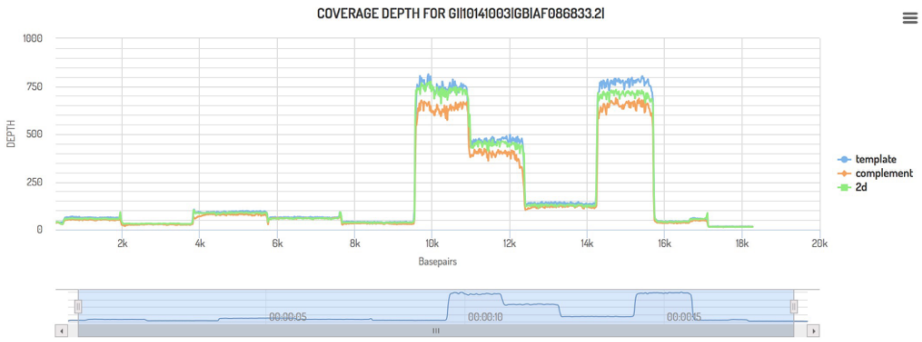


#7

Potential Applications

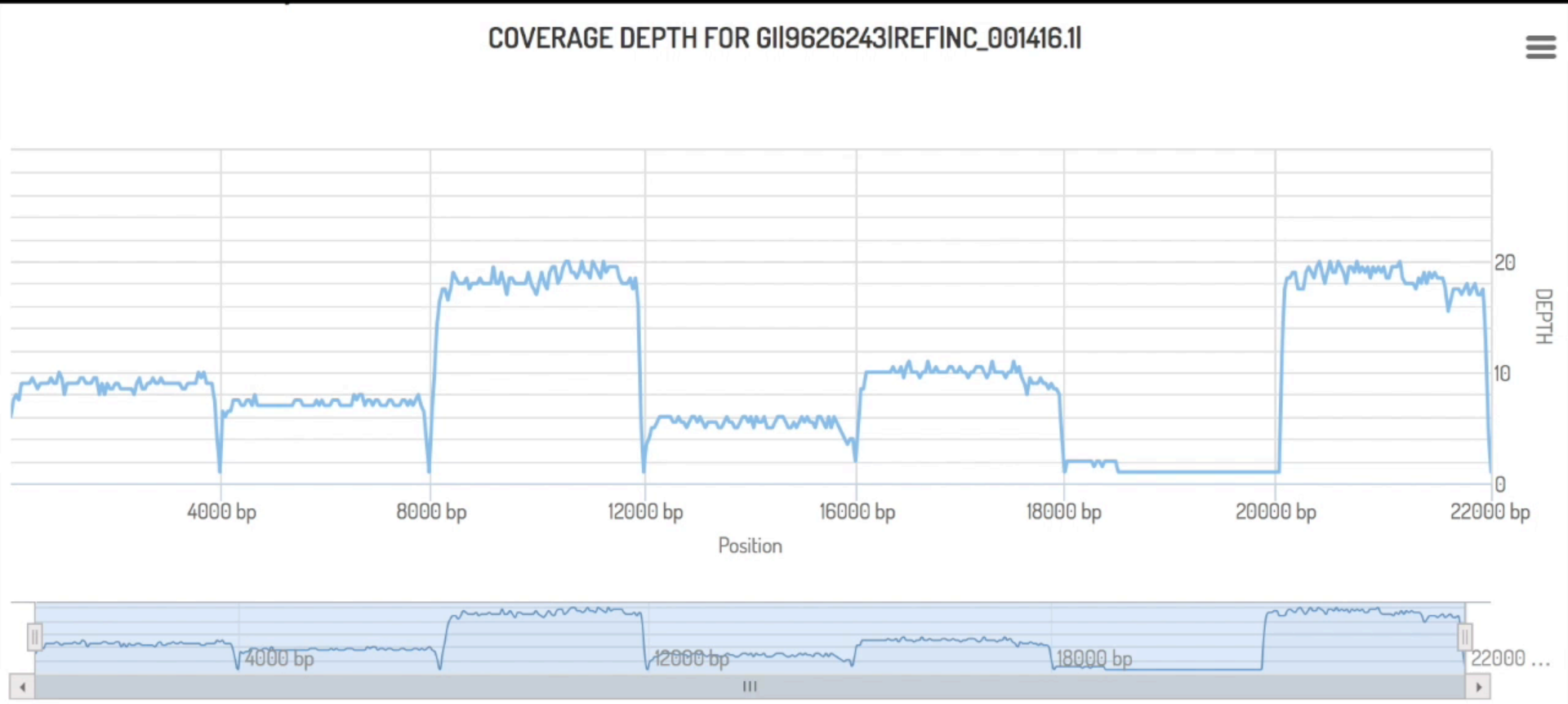
Applications for Squiggle Matching

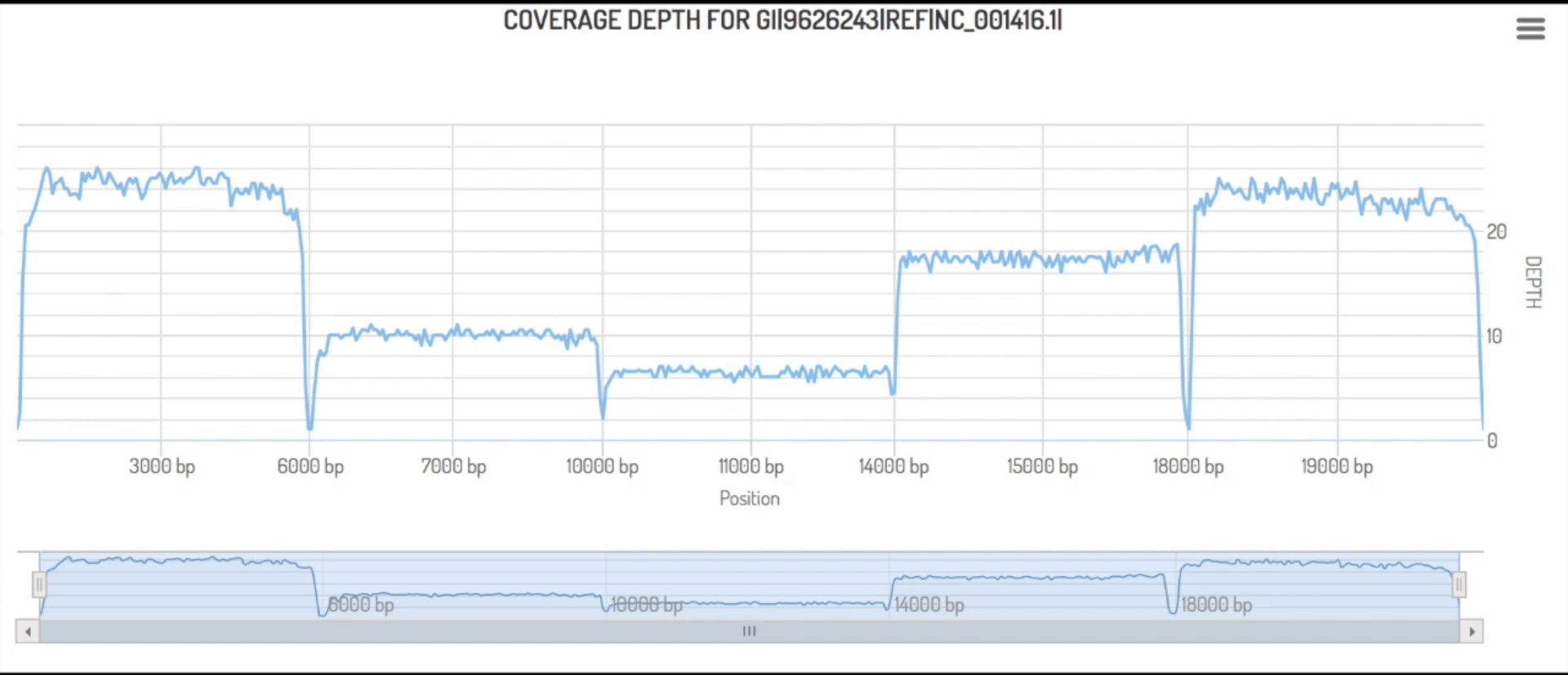
Amplicon Sequencing

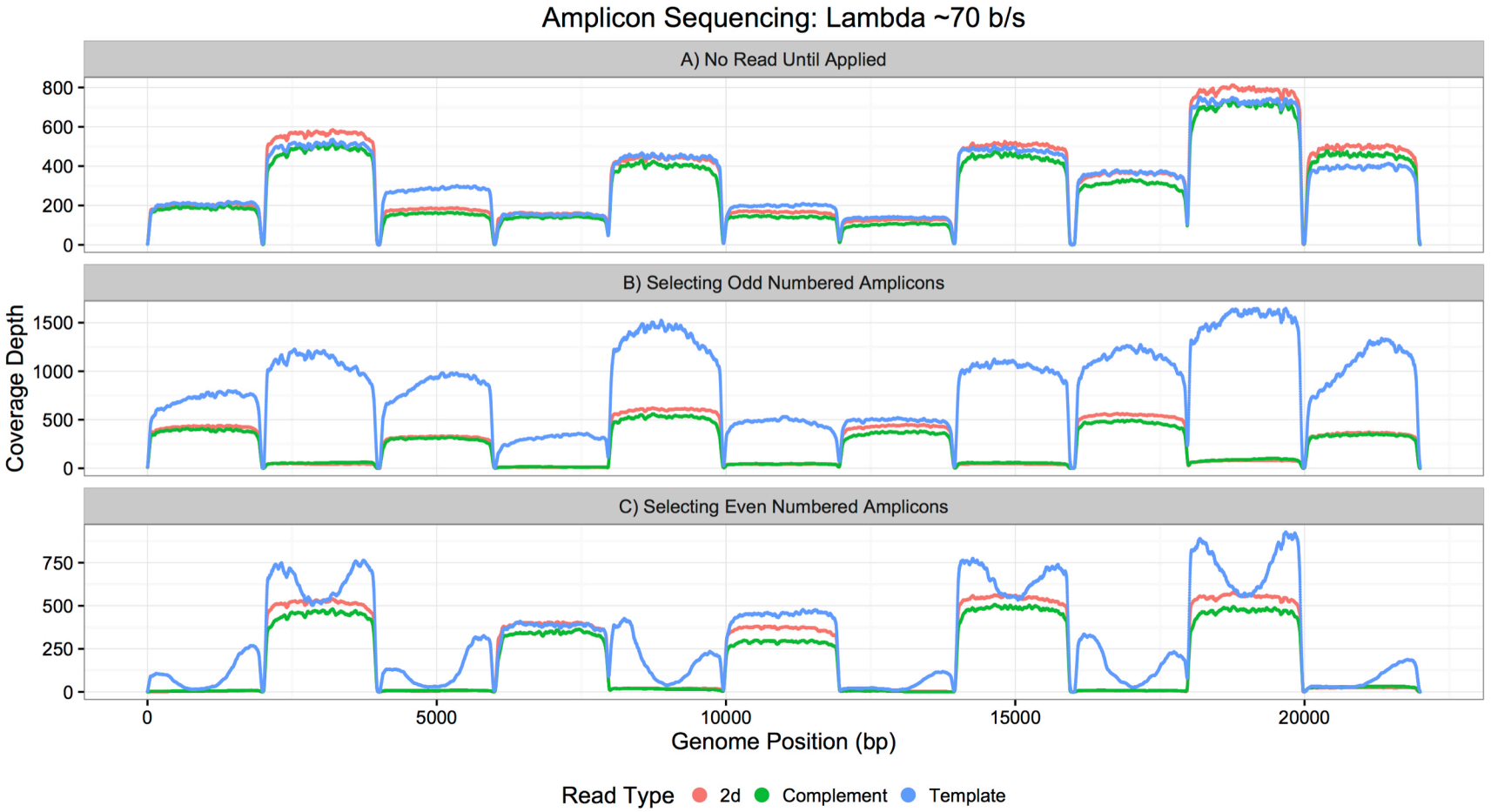


11 amplicons spanning 1 viral genome

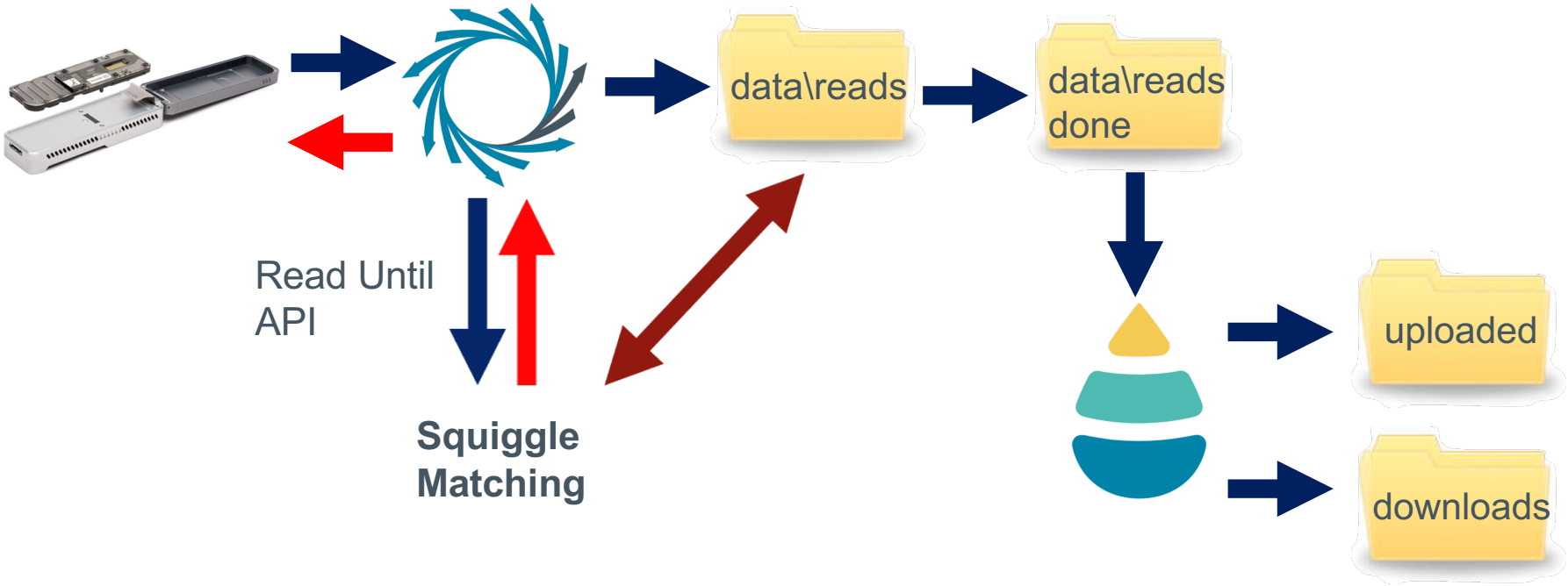
Aligned in basespace by BWA in minoTour

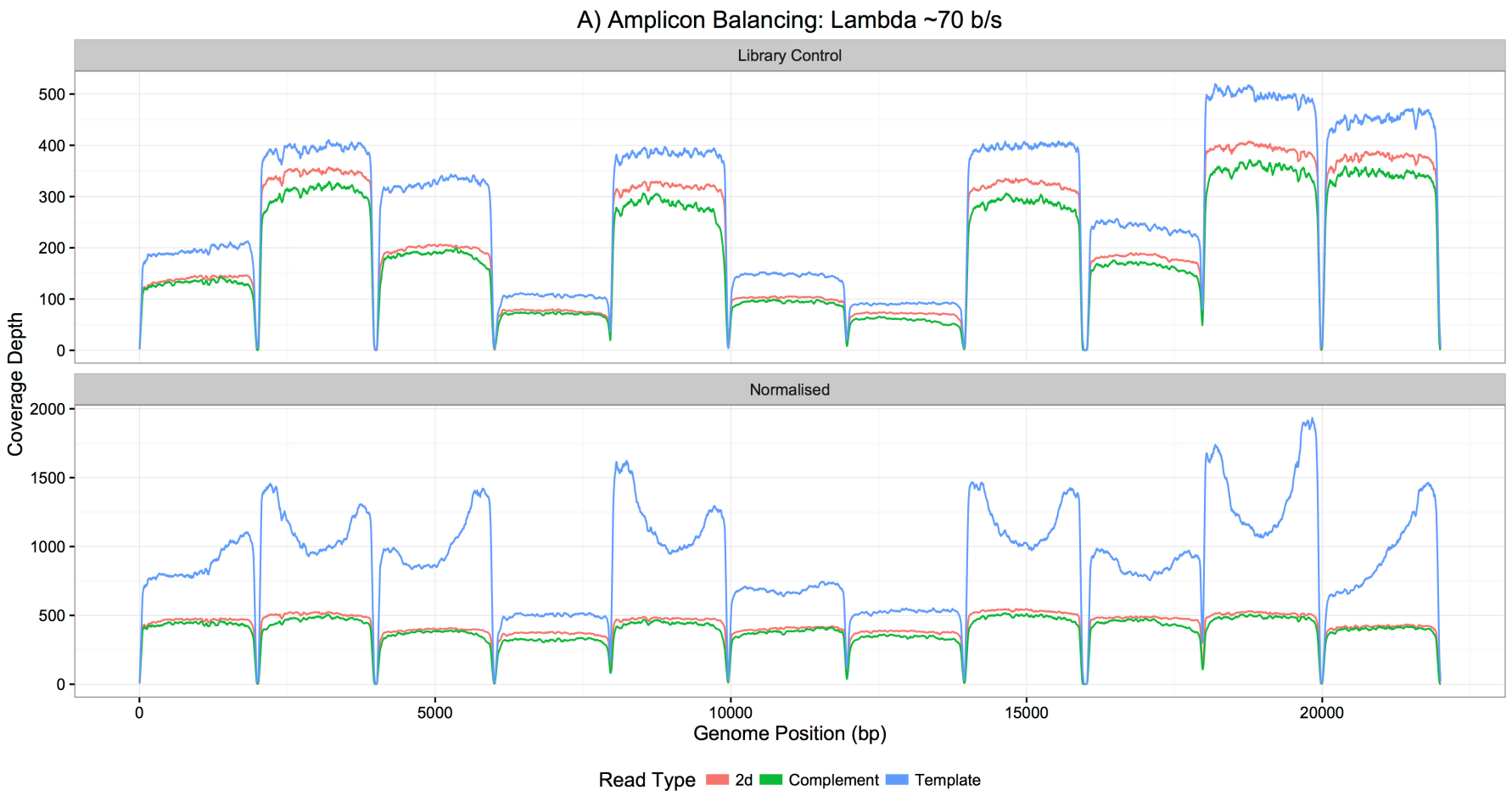

COVERAGE DEPTH FOR GI|9626243|REF|NC_001416.1|
20
10
0
DEPTH
4000 bp
8000 bp
12000 bp
16000 bp
18000 bp
20000 bp
22000 bp
Position
4000 bp
12000 bp
18000 bp
22000 ...


COVERAGE DEPTH FOR GI|9626243|REF|NC_001416.1|
20
10
0
DEPTH
3000 bp
6000 bp
7000 bp
10000 bp
11000 bp
14000 bp
15000 bp
18000 bp
19000 bp
Position
6000 bp
10000 bp
14000 bp
18000 bp


Amplicon Sequencing: Lambda ~70 b/s
A) No Read Until Applied
B) Selecting Odd Numbered Amplicons
C) Selecting Even Numbered Amplicons
Coverage Depth
800
600
400
200
0
1500
1000
500
0
750
500
250
0
0
5000
10000
15000
20000
Genome Position (bp)
Read Type
2d
Complement
Template

Read Until Workflow – Counting Amplicons




A) Amplicon Balancing: Lambda ~70 b/s
Library Control
Normalised
Coverage Depth
0
100
200
300
400
500
0
500
1000
1500
2000
0
5000
10000
15000
20000
Genome Position (bp)
Read Type
2d
Complement
Template

#8

Challenges

minoTour (v0.7)

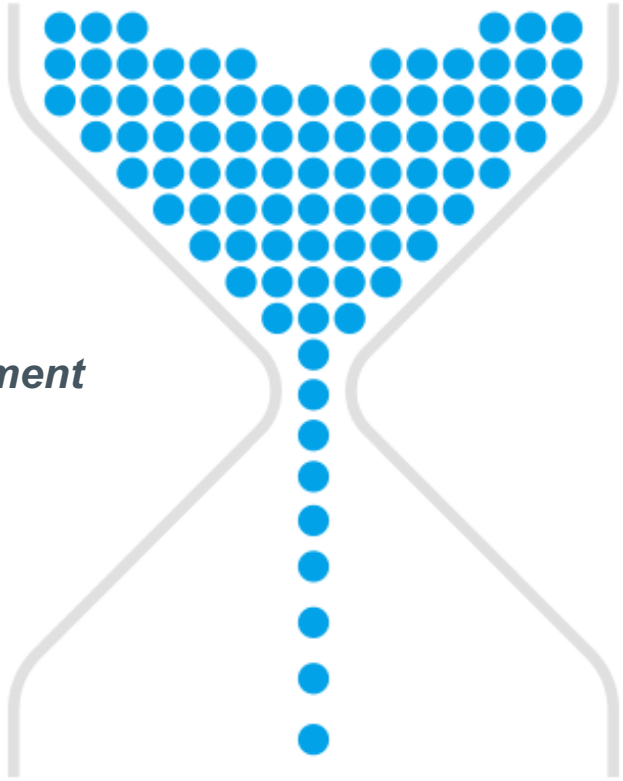
Challenges:

Yield

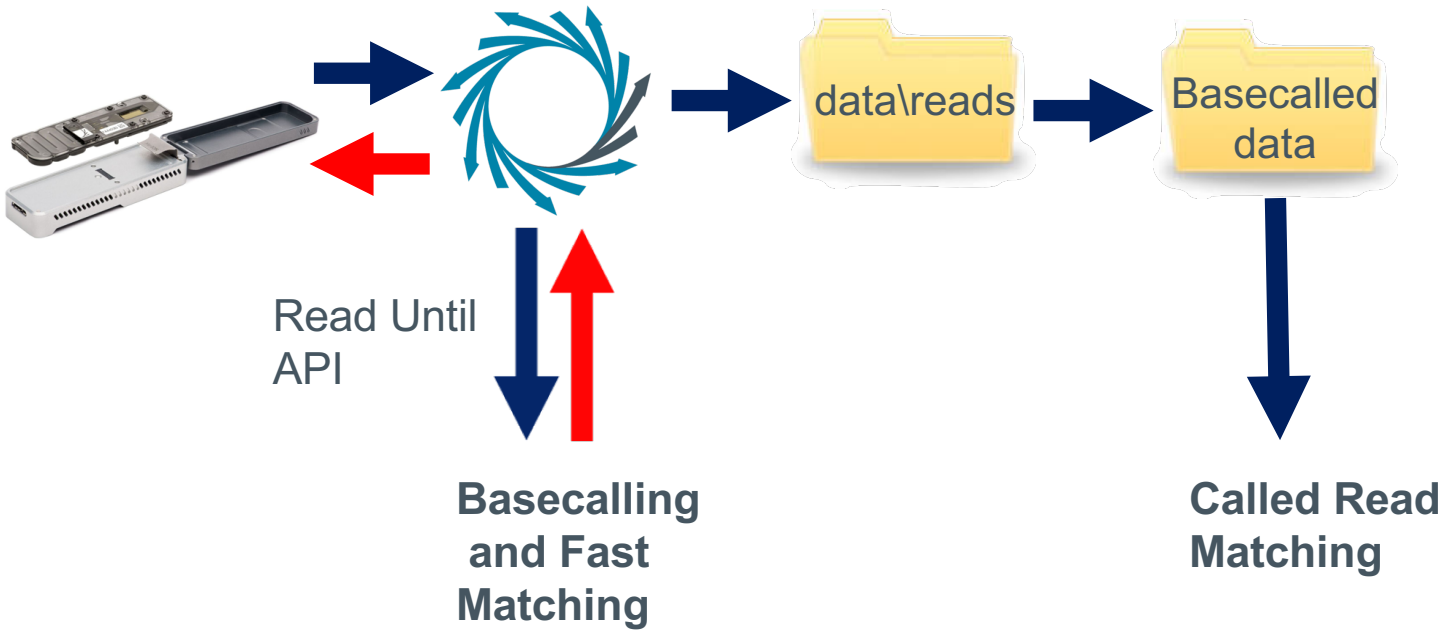
Rapid Asynchronous Read Alignment

Live Basecalling

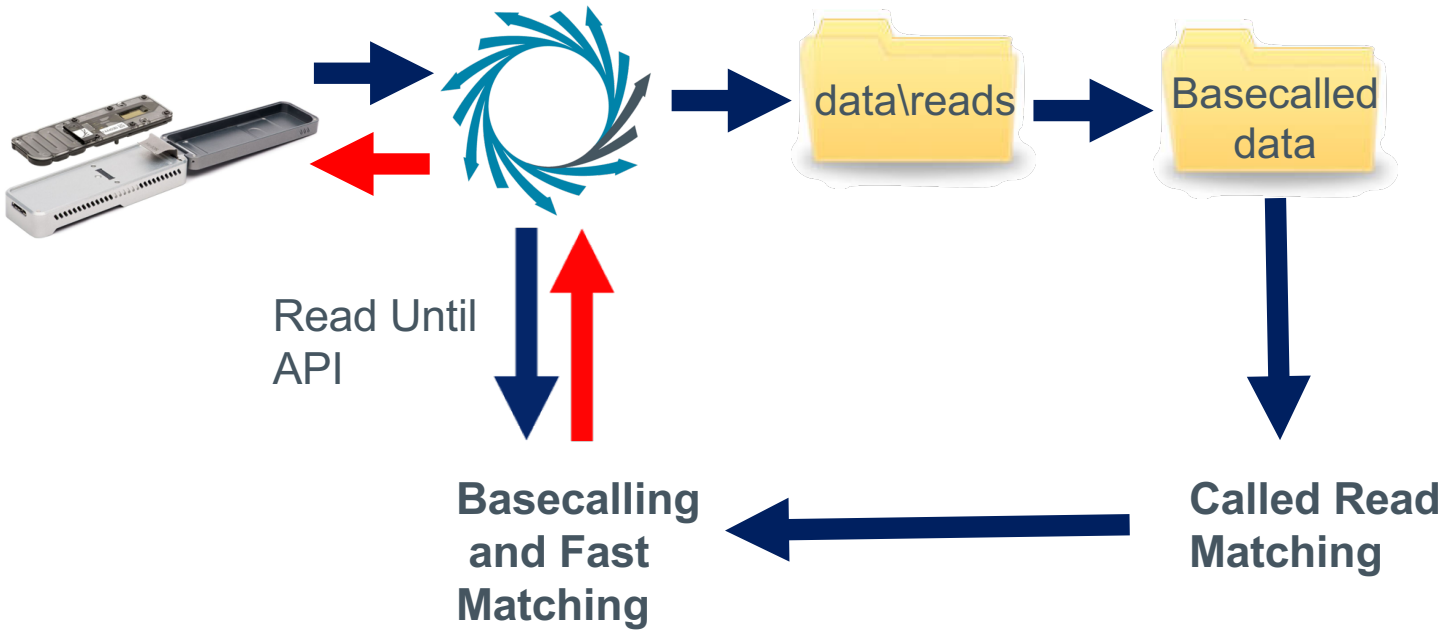
minoTour currently suited to small genomes (<150 mB)



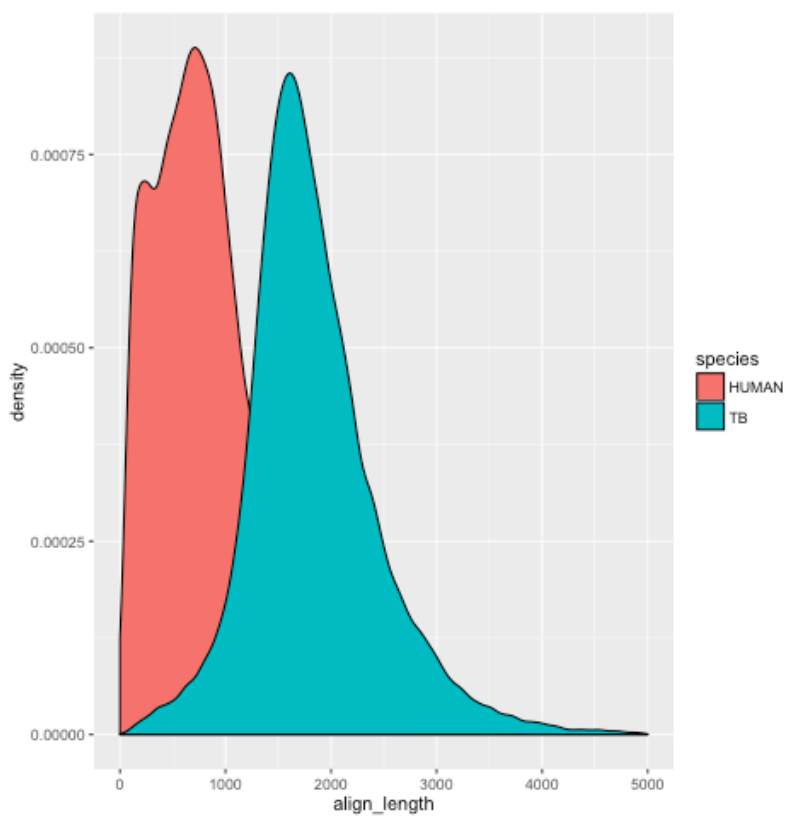
Read Until Workflows – Short and Long Loop

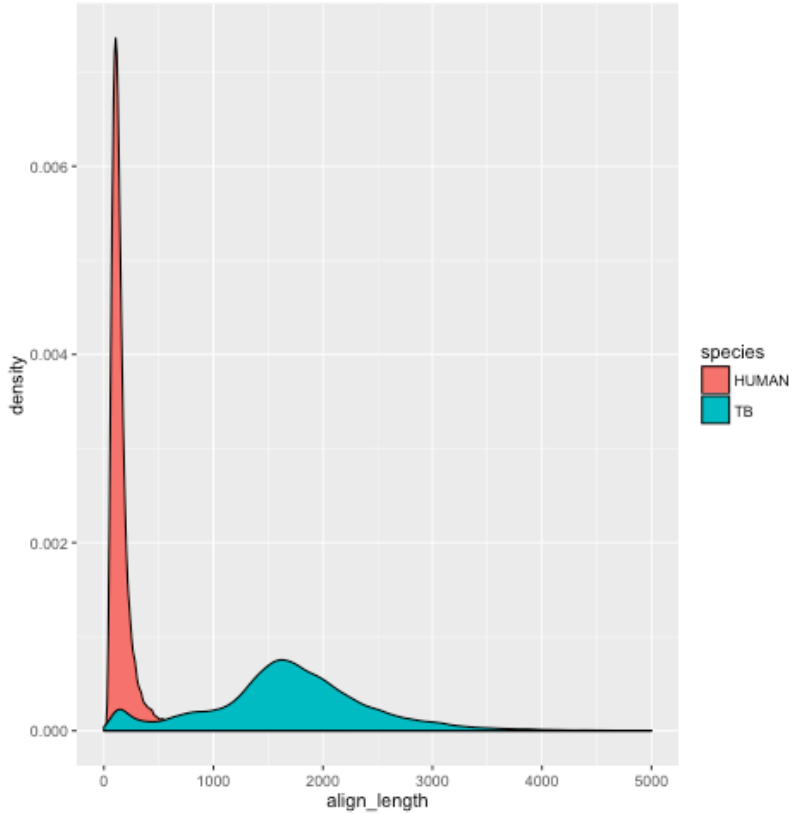


Read Until Workflows – Short and Long Loop

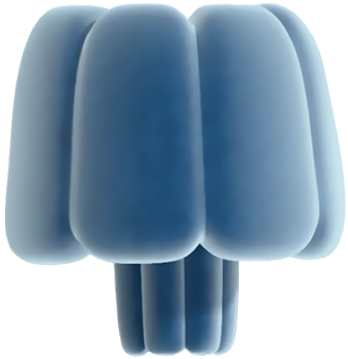


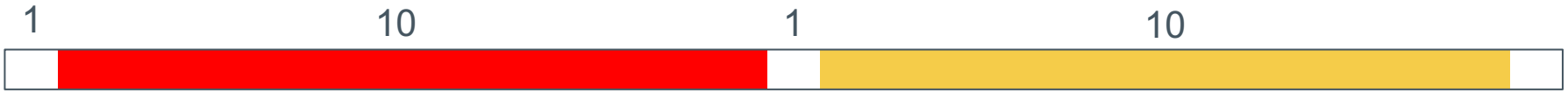
Alternative Workflows – rejecting host.





How does Read Until help?





Lower Bound ~ 400 bases
Allow 0.5 Seconds for match
Approximately 625 bases

Mean 10 kb library

Theoretical Max
Enrichment – 16 Fold

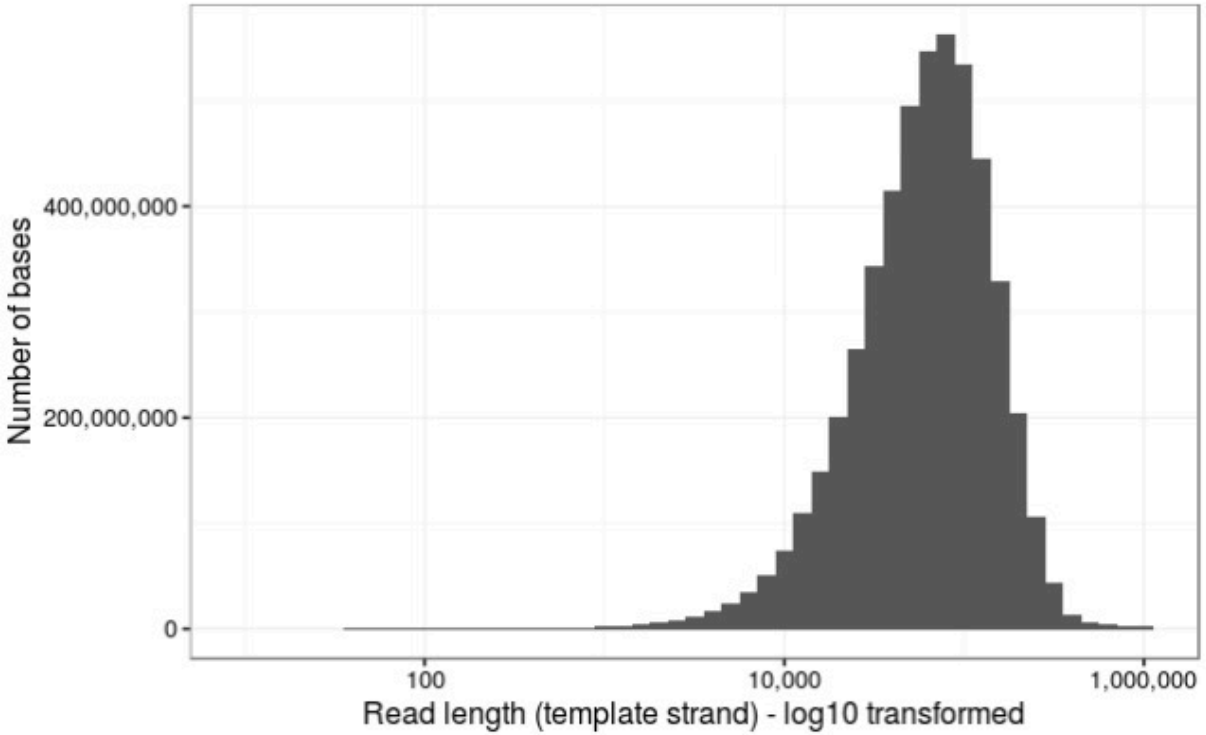
Pinned Tweet

Nick Loman @pathogenomenick · Mar 10

New blogpost: "Thar she blows! Ultra long read method for nanopore sequencing" lab.loman.net/2017/03/09/ult… full protocol included by @Scalene !

22 184 179

16x ≥ 50x enrichment



Alignment stats

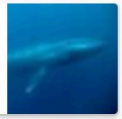
Wow! The longest 10 reads in this dataset are:

```
1113805 916705 790987 778219 771232 671130 646480 629747 614903 603565
```

!!!

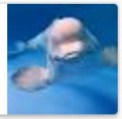
Whale › Mass

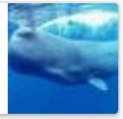
Killer whale: 3,600 – 5,400 kg 

Blue whale: 140,000 kg 

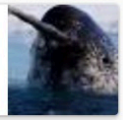
Short-finned pilot whale: 1,000 – 3,0… 

Humpback whale: 30,000 kg 

Beluga whale: 1,400 kg 

Sperm whale: 15,000 kg 

North Pacific right whale: 50,000 – 8… 

Narwhal: 940 kg 

North Atlantic right whale: 40,000 – … 

Human Genome Data Release

Link

http://github.com/nanopore-wgs-consortium/NA12878

13M reads from 29 flowcells (9M up with ~4M reads remaining to post)

>20x sequence coverage of NA12878 Genome

Contributors

Mark Akeson [1], Andrew D. Beggs [2], Thomas Nieto [2], Miten Jain [1], Nicholas J. Loman [3], Matt Loose [4], Sunir Malla [4], Justin O'Grady [5], Hugh E. Olsen [1], Josh Quick [3], Hollian Richardson [5], Jared T. Simpson [6,7], Terrance P. Snutch [8,9], Louise Tee [2], John R. Tyson [8,9]

1 University of California, Santa Cruz, Santa Cruz, CA, USA

2 University of Birmingham, Birmingham, B15 2TT

3 Institute of Microbiology and Infection, School of Biosciences, University of Birmingham, Birmingham, UK

4 DeepSeq, School of Life Sciences, University of Nottingham, Nottingham, UK

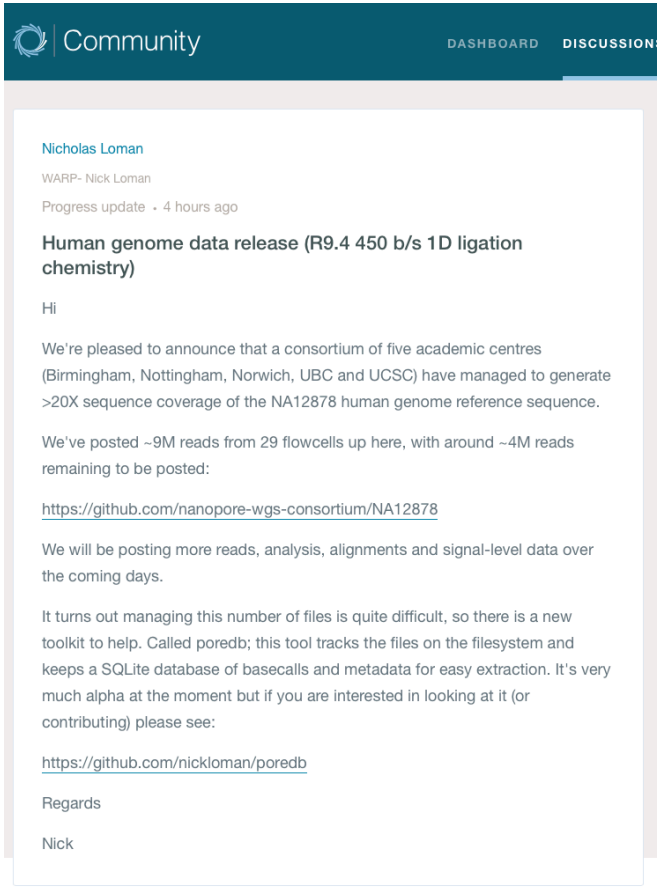
5 Norwich Medical School, University of East Anglia, Norwich, UK

6 Ontario Institute for Cancer Research, Toronto, Canada

7 Department of Computer Science, University of Toronto, Toronto, Canada

8 Michael Smith Laboratories, University of British Columbia, Vancouver, Canada

9 Djavad Mowafaghian Centre for Brain Health, University of British Columbia, Vancouver, Canada



Acknowledgements

minoTour.nottingham.ac.uk
github.com/minoTour/minoTour

University of Nottingham: Martin Blythe, Sunir Malla, Mike Stout, Teri Evans
University of Birmingham: Nick Loman, Josh Quick

Nanopore WGS Consortium: github.com/nanopore-wgs-consortium/NA12878

EBI: Ewan Birney, Guy Cochrane

Oxford: Zam Iqbal

@mattloose matt.loose@nottingham.ac.uk

minoTour

The University of Nottingham
UNITED KINGDOM · CHINA · MALAYSIA

amazon web services

BBSRC
bioscience for the future

Oxford NANOPORE Technologies