# Sequencing versus genotyping

Genotyping

- Determining the sequence of an individual at a series of specific loci
- Determines a series of SNPs
- Doesn't capture rearrangements, unknown/unprobed changes
- Cheaper to collect and analyze data
- Sufficient for many genomics questions

Sequencing

- Determines identity of each base in the genome/ library (hopefully)
- Captures rearrangements, known and unknown SNPs
- Needs large datasets
  - many fold more for coverage
- Lots more data to analyze
- May provide more data than is needed for a specific question

# When is genotyping better?

- Known series of DNA SNPs you care about
  - Cancer panel for determining cancer treatment
  - Ancestry analysis (Ancestry.com/23andMe)
  - Disease diagnosis/treatment
- Known set of RNAs you care about
  - Cancer diagnosis
  - Metabolic analysis
- Narrow scope of search
  - What genes/regions of genome might be correlated to a disease

# When is sequencing better?

- Not sure what you're looking for
  - Are there changes in the genome, other than SNPs, that affect a phenotype
  - Metagenomics
  - Looking for new/different splicing, gene expression, or noncoding RNAs
- Genome construction

# What are the problems with 2nd generation sequencers?

# What are the problems with 2nd generation sequencers?

- Short reads (200-600bp)
- Amplification errors
- Time/expense

GATCTTCGTACTGAGT CTG
GATCTTCGTACTGAGT CTG

GATTTTCGTACGGAAT TGA
GATTTTCGTACTGAGT TTG

GATCTTCGTACTGAAT CTA
GATTTTCGTACGGAAT TGA

GATTTTCGTACTGAAT TTA
GATCTTCGTACGGAAT CGA

Remember haplotypes?

Often large distances between loci (1-2kb). Short reads make it hard to identify these.

How can we fix?

# Solve Haplotype problem

1. Mate Pairs
   - If we get lucky see loci of interest in mate pair library
2. Get long reads
   - Ideally 1-2kb

3rd generation sequencers try to solve this problem

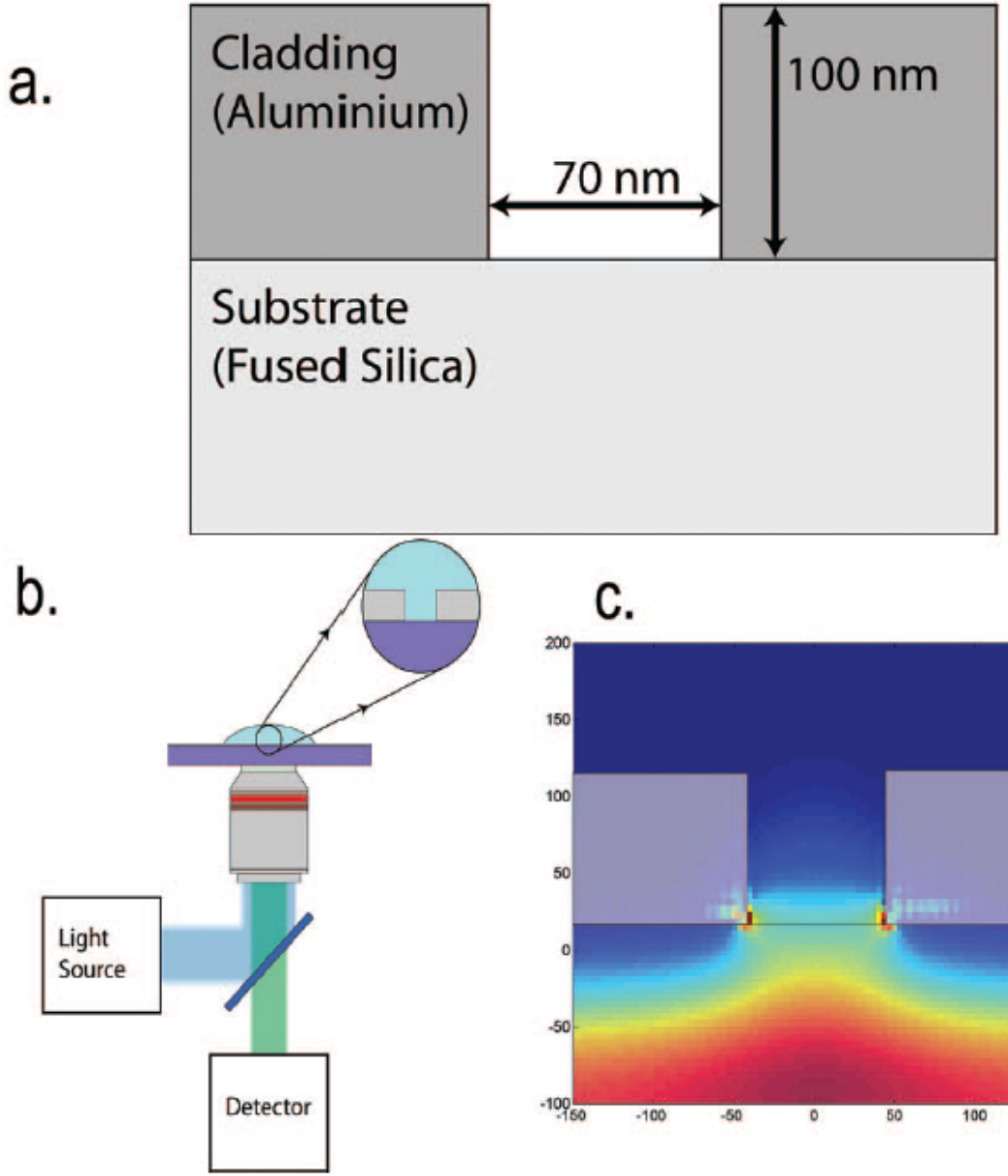
# 3rd generation sequences

- Single Molecules

- Three main platforms
  - Helicos (short reads, died in ~2 years)
  - Pacific Biosciences
  - Oxford Nanopore

# Pacific Biosciences!

- Single-molecule, long reads!

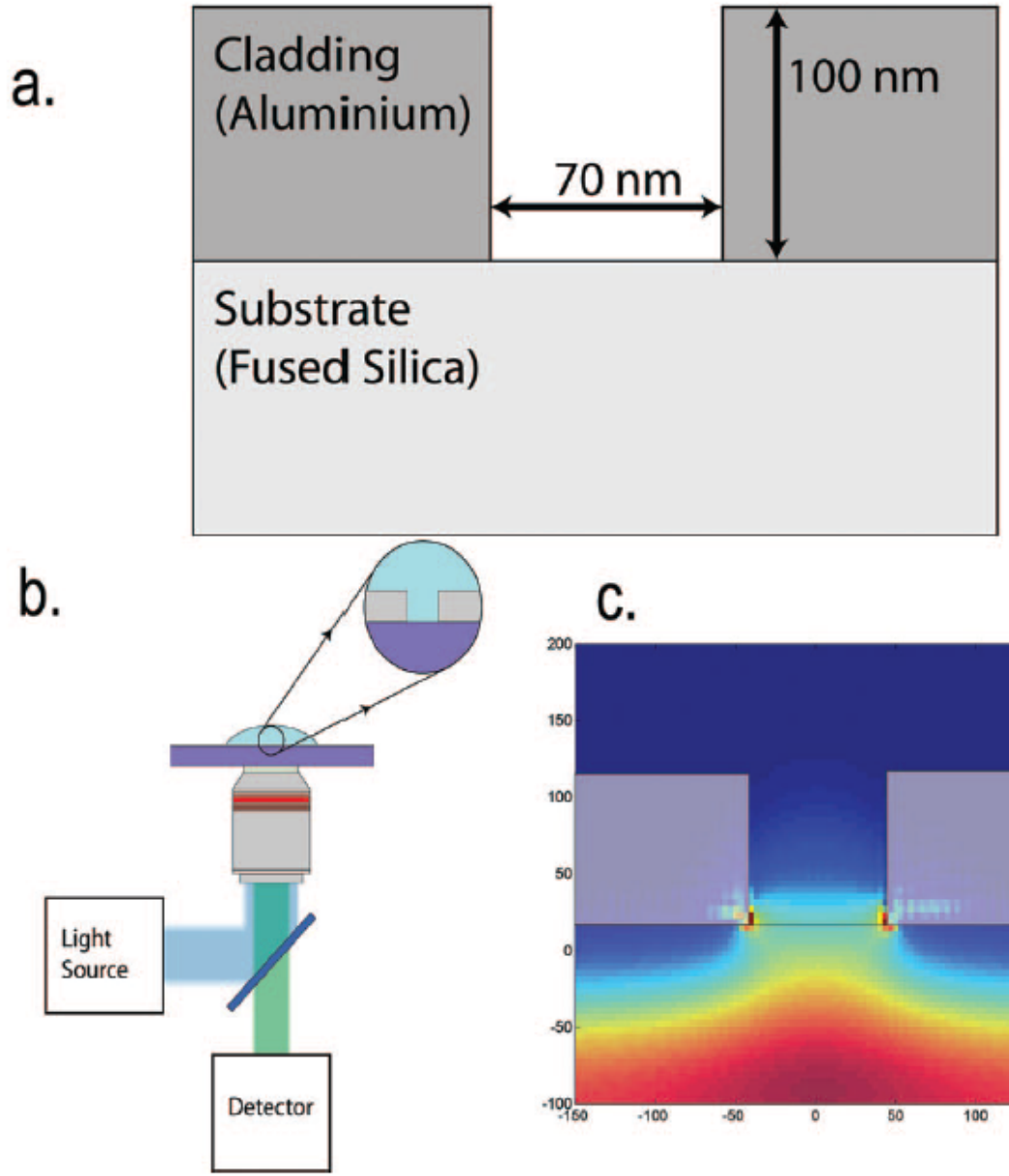



Zero-mode Waveguide

What are wavelengths of visible light?

380-750nm

Foquet, M., et al.; *Journal of Applied Physics* **103** , 034301 (2008); doi: 10.1063/1.2831366

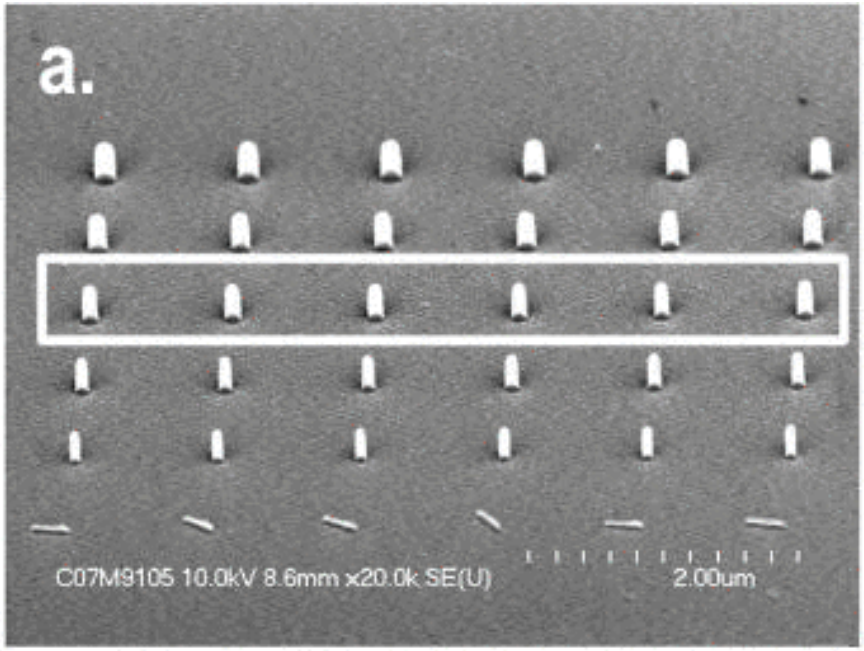


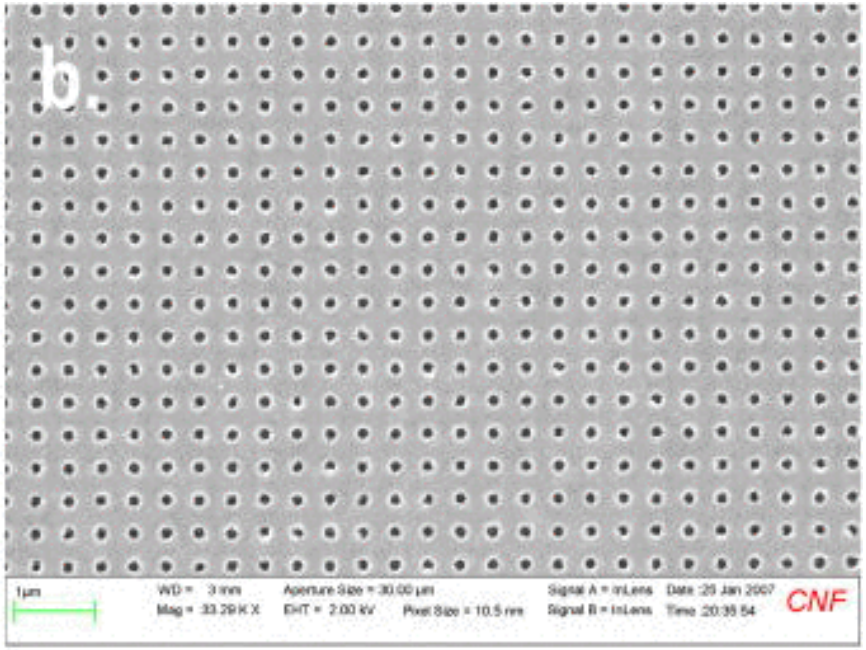

# Zero-mode Waveguide

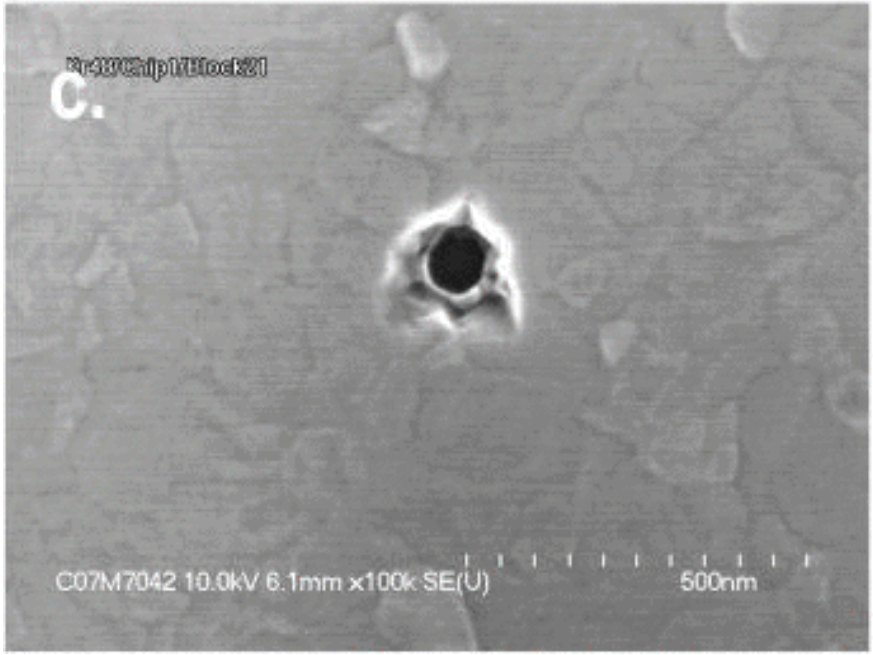
Only bottom 20-30 nm become illuminated

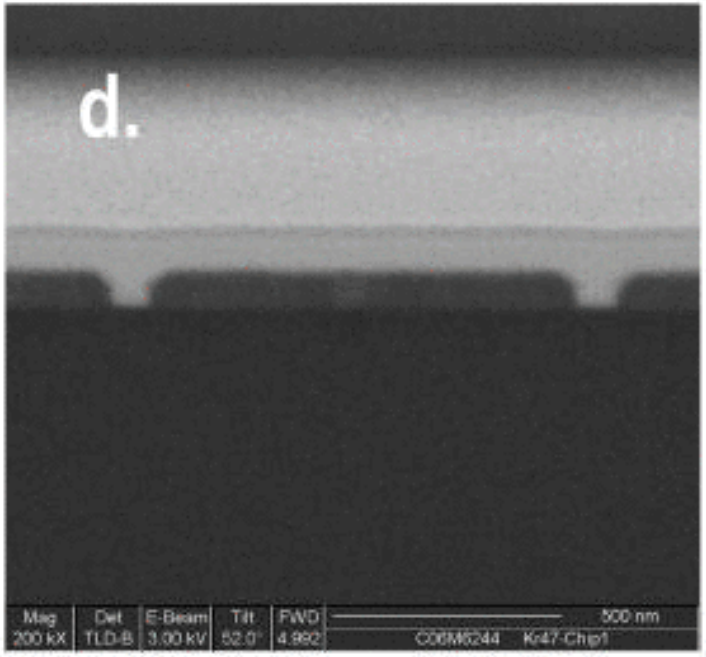
Makes visible only about 20 zeptoliters ($10^{-21}$ liters) of volume

Foquet, M., et al.; *Journal of Applied Physics* **103** , 034301 (2008); doi: 10.1063/1.2831366

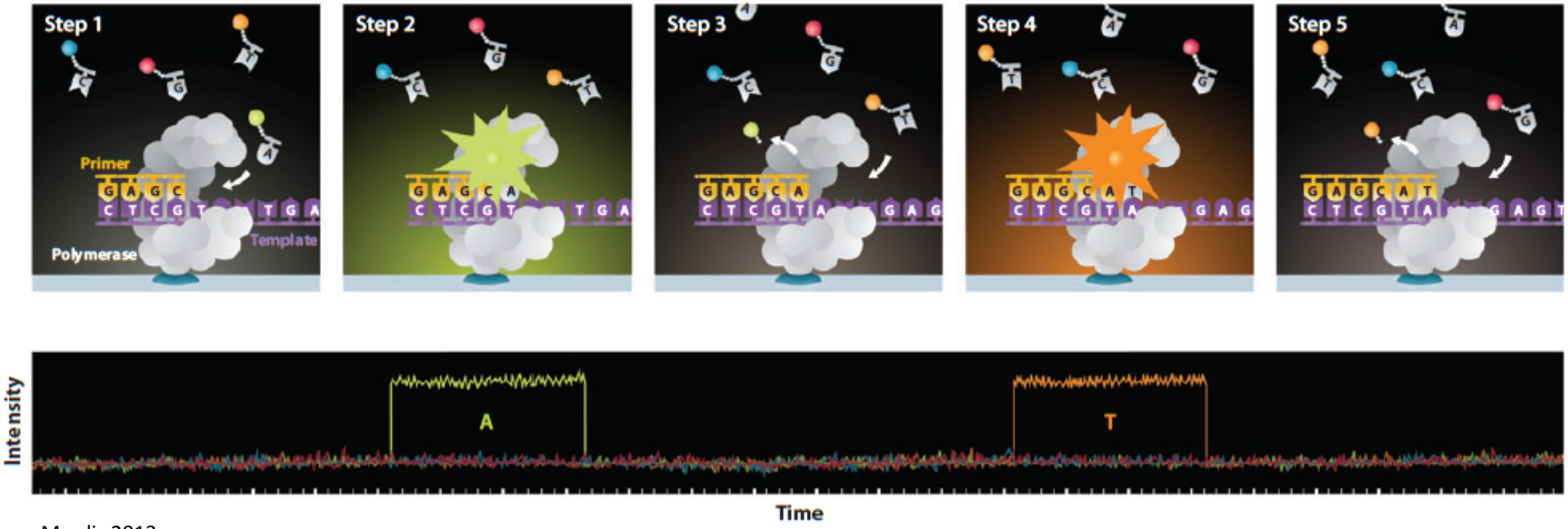


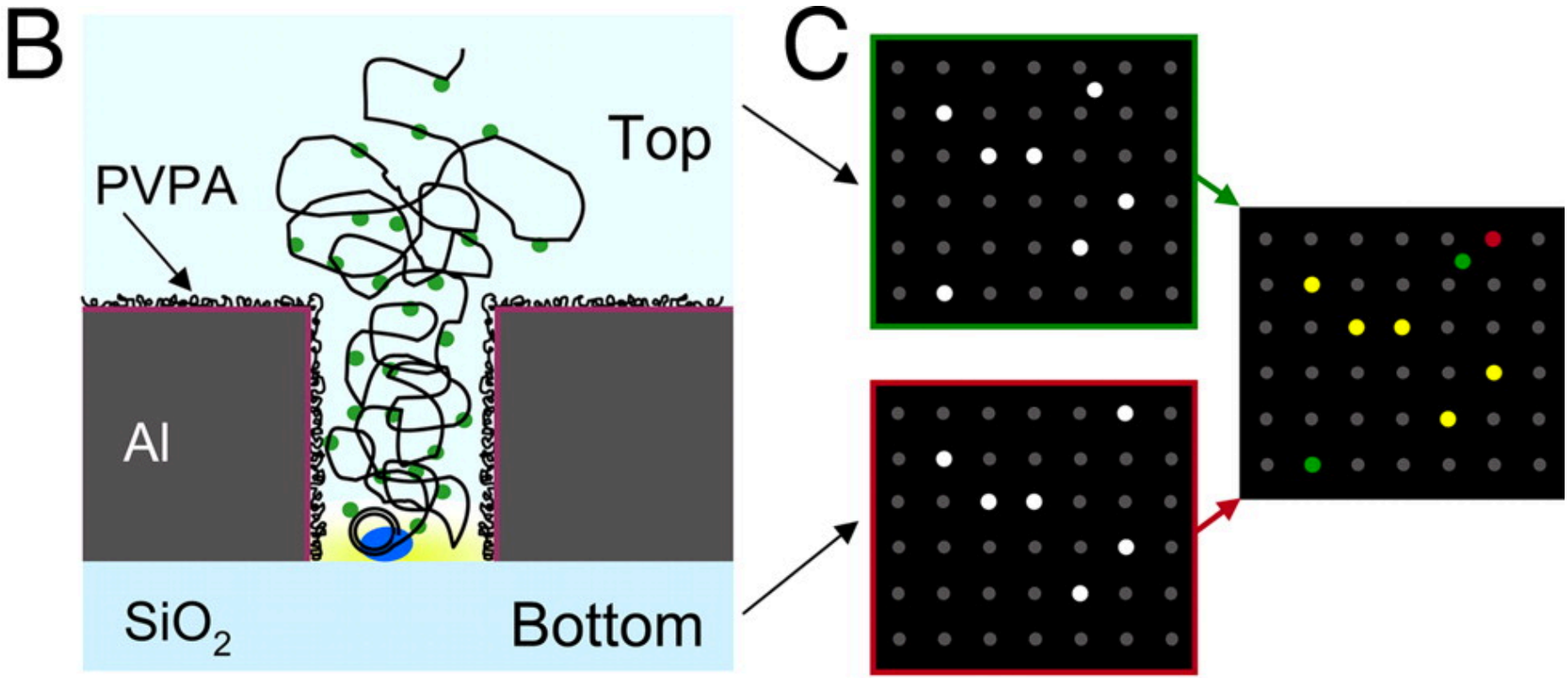




Foquet, M., et al.; *Journal of Applied Physics* **103** , 034301 (2008); doi: 10.1063/1.2831366

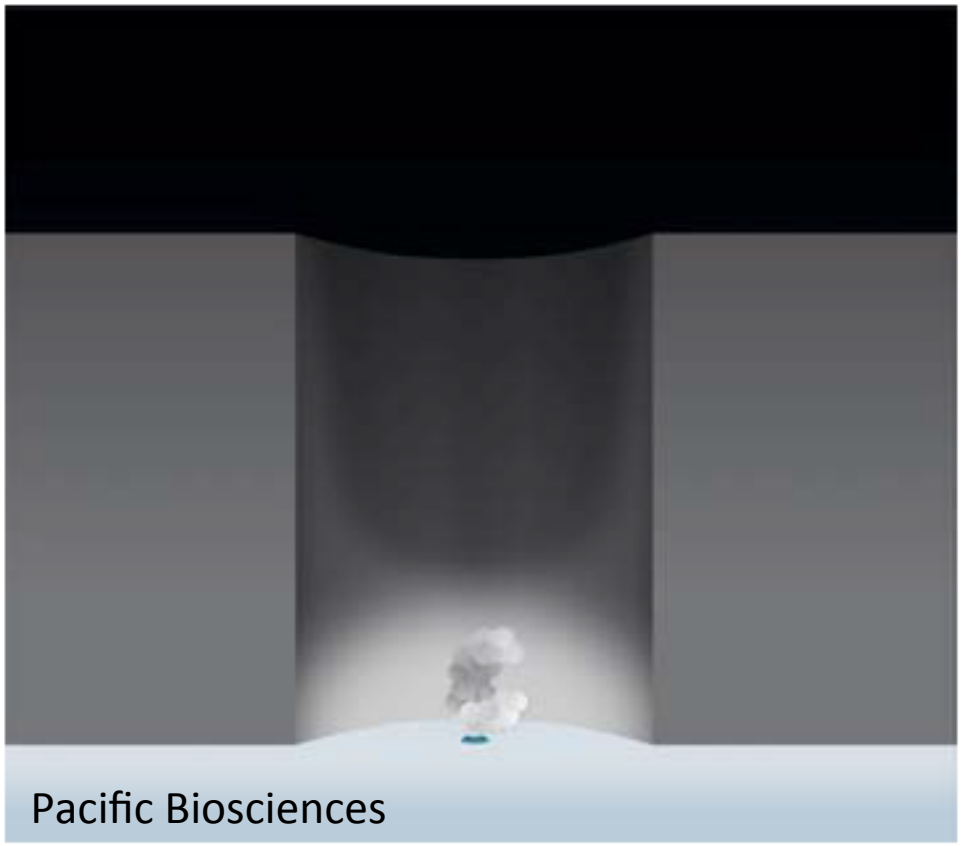

Step 1
Step 2
Step 3
Step 4
Step 5
Primer
Polymerase
Template
Intensity
A
T
Time

Mardis 2013

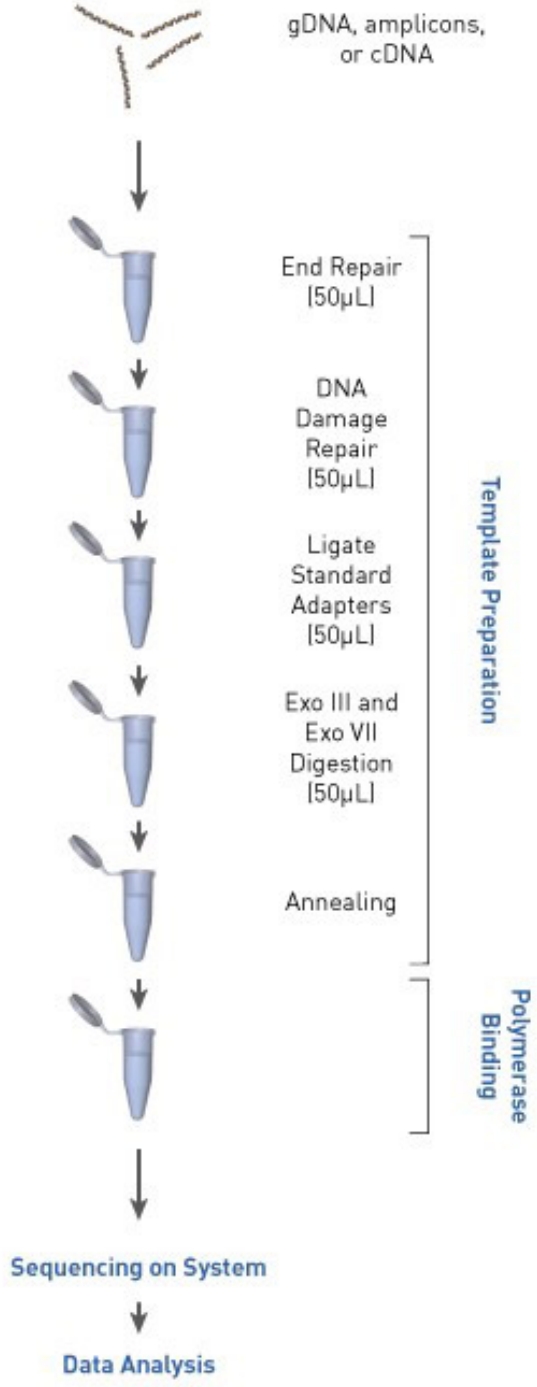



Korlach, J., et al.; PNAS,doi: 10.1073/pnas.0710982105

Need to make sure polymerase binds to bottom of ZMW

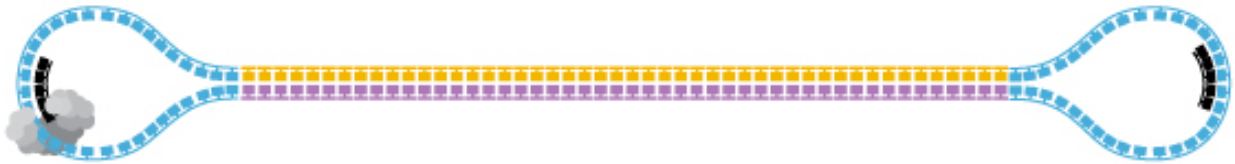


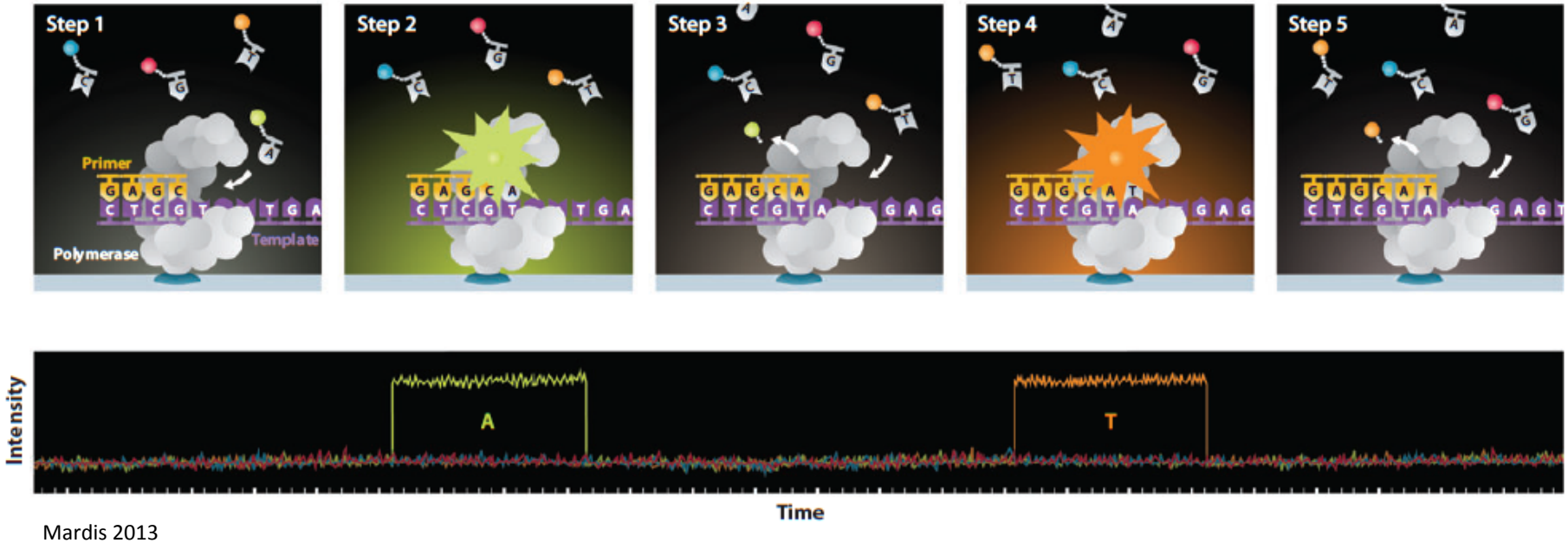

Need polymerase to bind in an active form:
Proprietary technology
Likely some adaptation of biotin/streptavidin or another technology of this type




Start with fragmented DNA

- Large fragment 500bp-20kb
- Ligate adapters (loops)
- Bind polymerase and primer



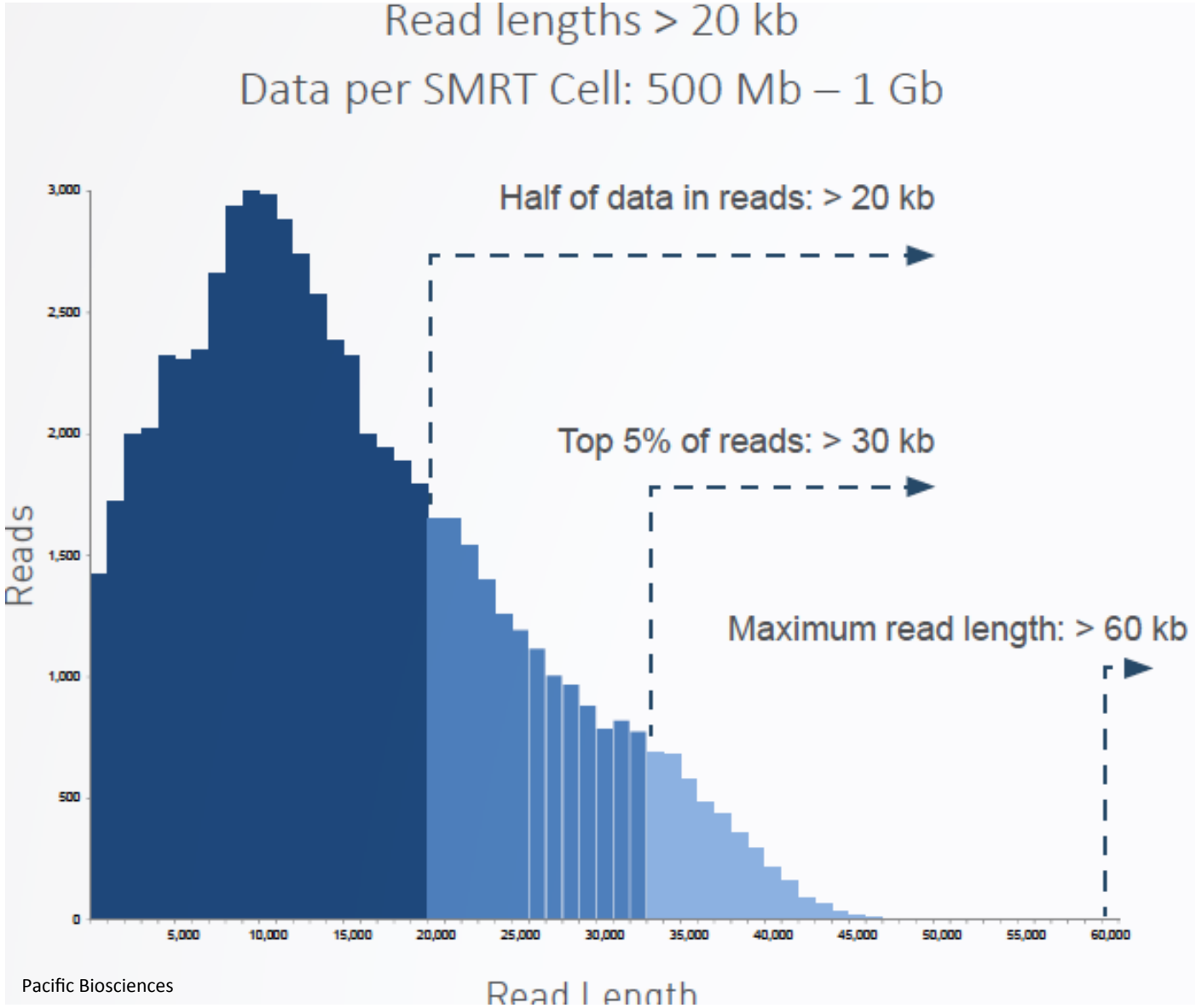



Mardis 2013

Need to change the speed of polymerase:
Faster or slower?

# Accuracy

- Single-pass
  - ~86%
  - 30x passes ~99.999%
- If nucleotide take an unusually short or long time to incorporate (dwell time), often miscount that base – leading to indel errors




Most reads are >20kb

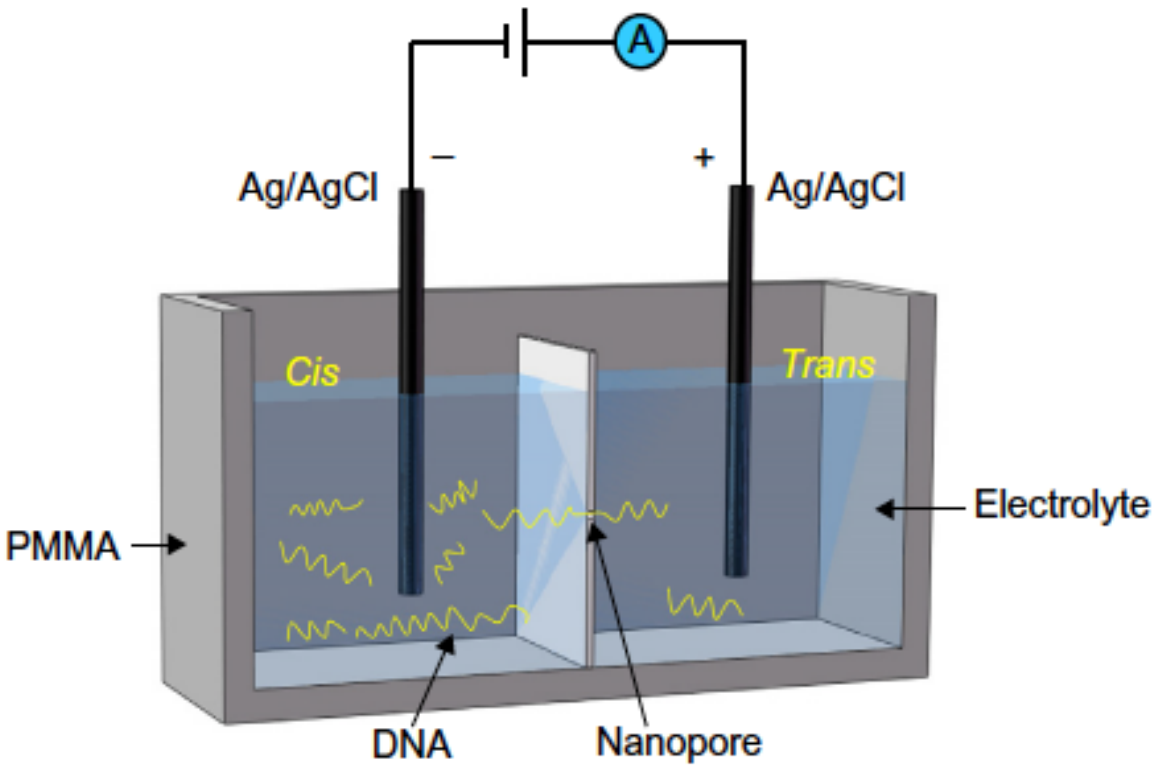
# PacBio Specs

- PacBio RSII
  - Each cell 500Mb to 1Gb
    - Up to 16 cells run automatically
  - 10-60 kb per read
  - 50k reads
  - Up to 4 hours
- PacBio Sequel
  - Each cell 5Gb-10Gb
    - Up to 16 cells run automatically
  - 10-60 kb per read
  - 500k reads
  - Up to 4 hours

# Questions?

# Oxford Nanopore

- Newest sequencer on the market

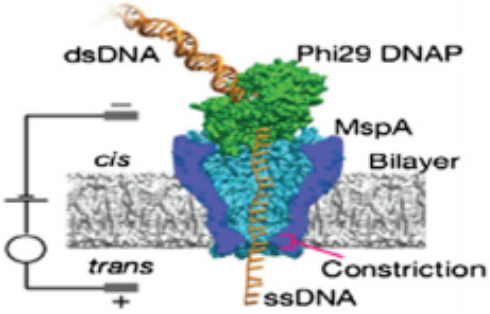


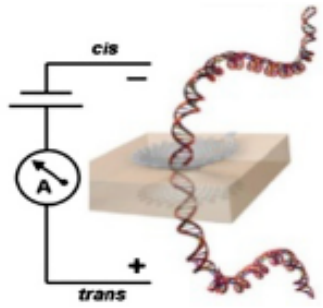

Feng,Y. et al.; *Genomics, Proteomics & Bioinformatics* **13**(1), February 2015; http://doi.org/10.1016/j.gpb.2015.01.009.

- Measure changes in voltage as nucleic acids cross through nanopore
- Distinguish between each base

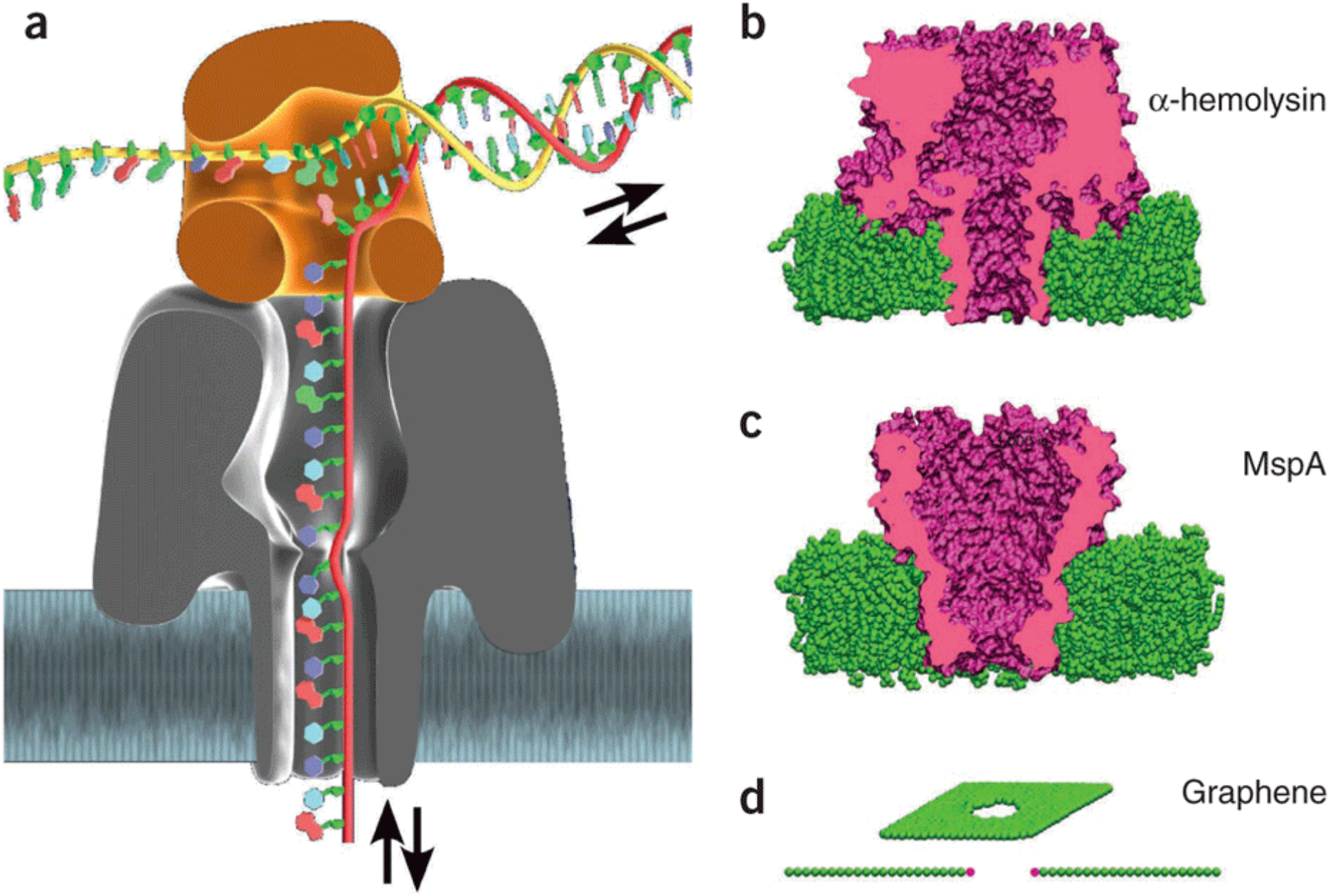
# Two main types

- Biological
  - Membrane as barrier
  - Protein provides nanopore
- Solid-state
  - Non biological barrier
  - Silicon and $Al_2O_3$ common
  - Protein may still be involved in providing motion

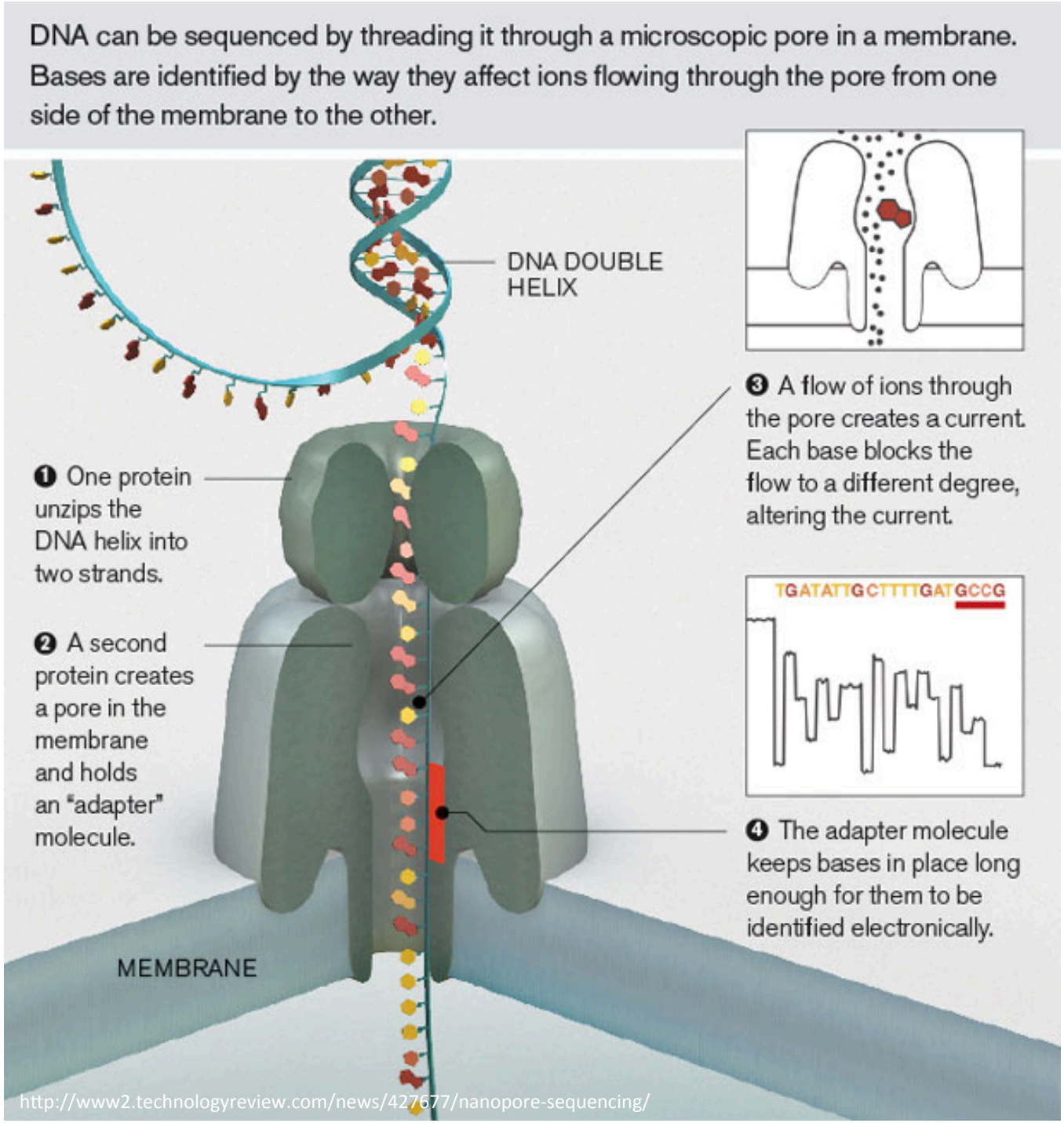



Feng,Y. et al.; *Genomics, Proteomics & Bioinformatics* **13**(1), February 2015; http://doi.org/10.1016/j.gpb.2015.01.009.

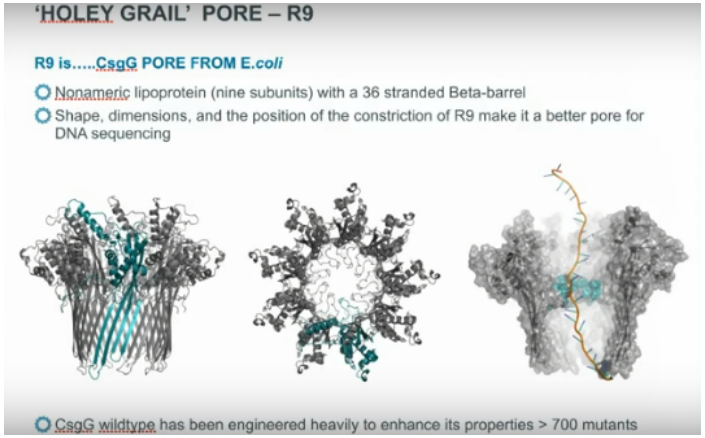



Grégory F Schneider& Cees Dekker; *Nature Biotechnology* **30**, 326–328 (2012) doi:10.1038/nbt.2181


DNA can be sequenced by threading it through a microscopic pore in a membrane. Bases are identified by the way they affect ions flowing through the pore from one side of the membrane to the other.
DNA DOUBLE HELIX
❶ One protein unzips the DNA helix into two strands.
❷ A second protein creates a pore in the membrane and holds an "adapter" molecule.
❸ A flow of ions through the pore creates a current. Each base blocks the flow to a different degree, altering the current.
TGATATTGCTTTTGATGCCG
❹ The adapter molecule keeps bases in place long enough for them to be identified electronically.
MEMBRANE
http://www2.technologyreview.com/news/427677/nanopore-sequencing/

# MANY lawsuits

- Illumina was suing Oxford over its use of MspA
  - Illumina licensed the technology developed here at UW, though they have not yet commercialized
- In August, they settled the suit as Oxford has moved to using a different molecule CsgC




http://omicsomics.blogspot.com/2016/03/oxfords-no-thanks-ive-already-got-one.html

# MANY lawsuits

- Illumina was suing Oxford over its use of MspA
  - Illumina licensed the technology developed here at UW, though they never commercialized it
- In August, they settled the suit as Oxford has moved to using a different molecule CsgC
- In November PacBio filed suit again. Infringing patent of using linked double stranded templates
  - Still ongoing

# MinION

Small (pocket sized)
Runs on USB drive
512 nanopore channels
Simple sample prep
No amplification needed
$1000 to buy and run twice

Oxford Nanopore

# 1D and 2D reads

- Need to fragment DNA, but less
  - Can't sequence whole chromosomes yet
- Need to add adapter to 1 end that will allow it to bind to protein nanopore
- Can either sequence 1 strand only or the forward and reverse strand
  - To sequence both (2D) ligate hairpin to one end, so when you reach the end, the other strand will come through backwards

# Oxford Nanopore Stats

MinION

- 5kb-200kb reads
- How much data depends on size of reads
- Nearly immediate access to data
  - can watch as it reads sequences
- $99 for sequence after purchased device
- Can be used outside of a lab

GridION5X

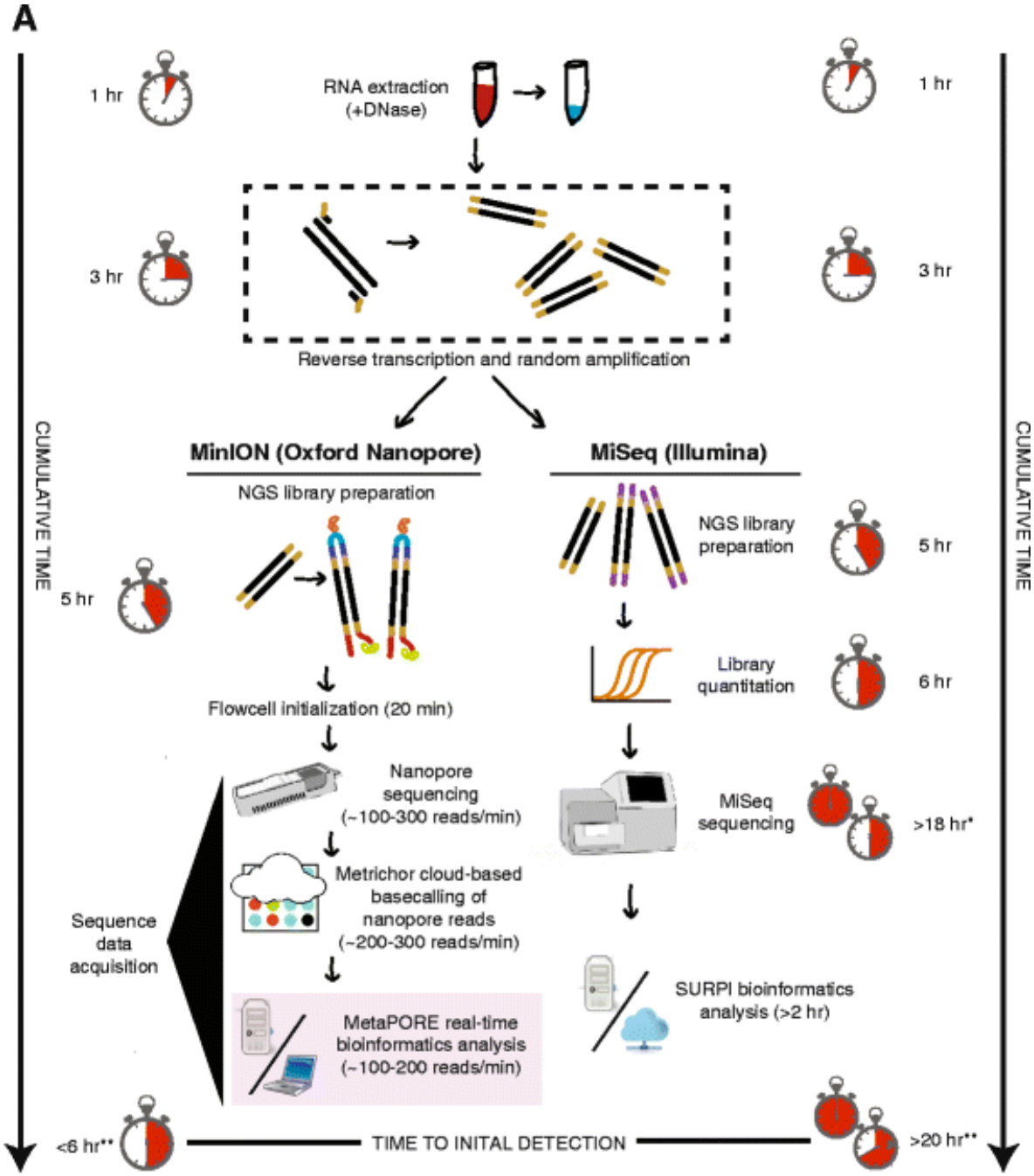
- 5 MinIONs in a box that can run then individually or together

PromethION

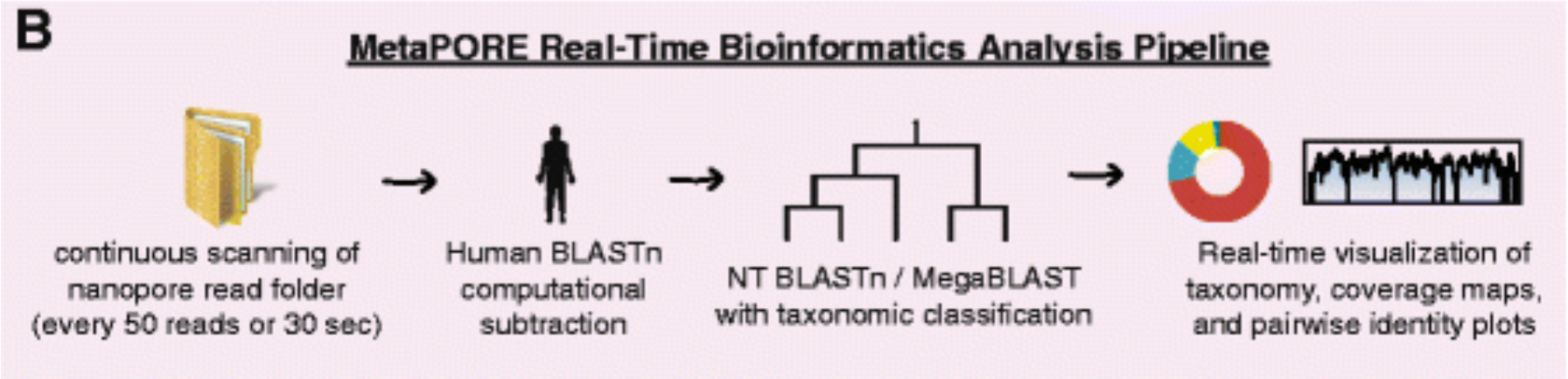
- Benchtop sequencer
- 48 MinION flow cells
- Computation system included in box
- Not on market yet, but in beta testing in labs

# Accuracy Problems

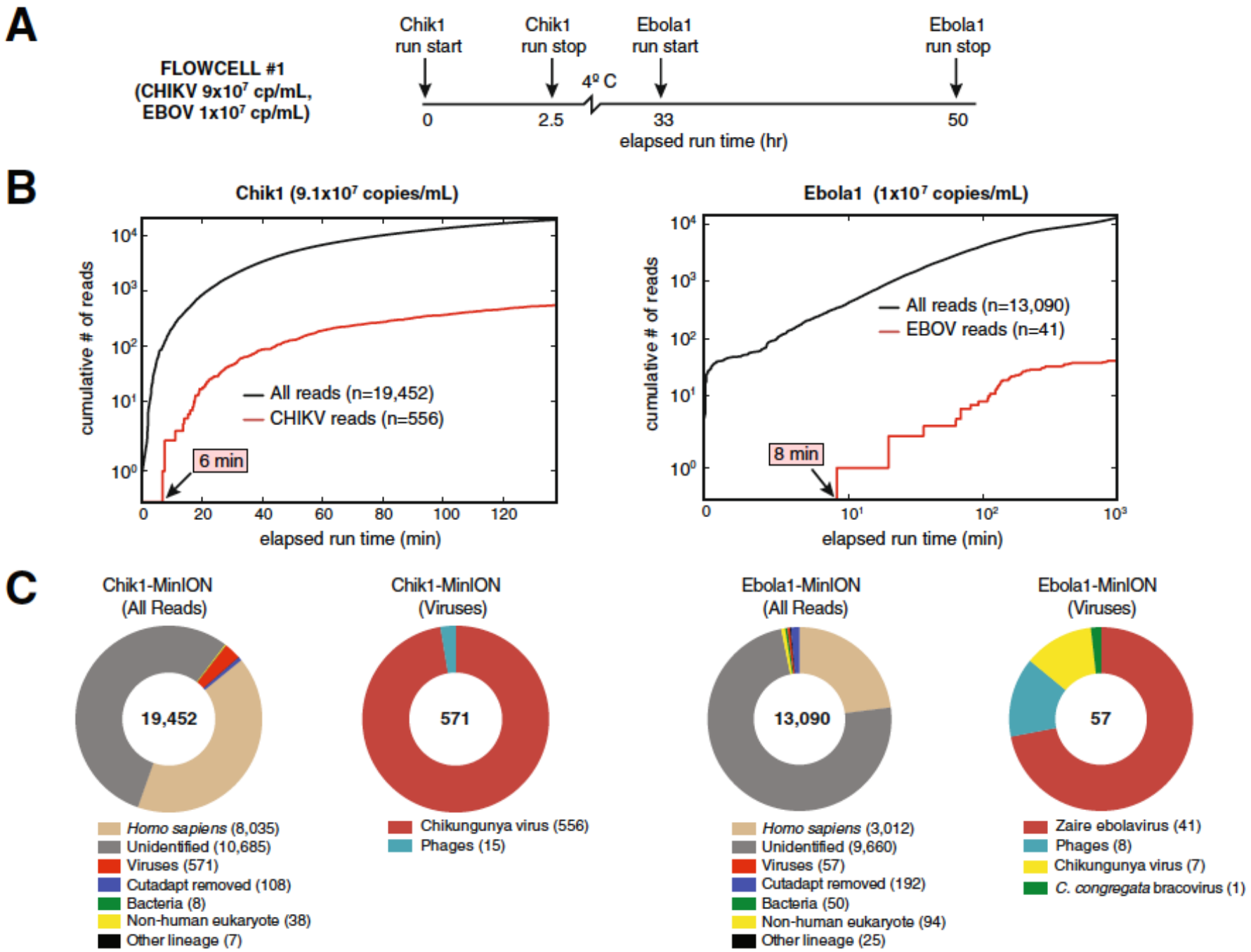
- Per base accuracy isn't great (like PacBio)
- 24% base call errors
- Unlike PacBio, can't reread enough to average out errors
- Ok for some applications



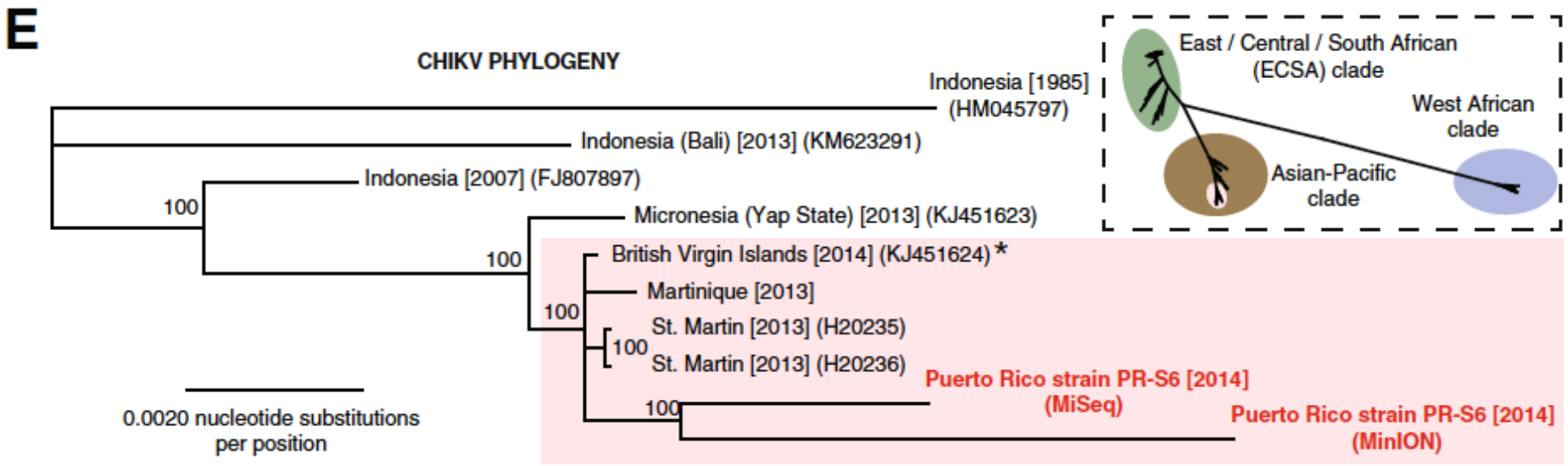

Greninger, AL, et al.l;*Genome Medicine* 2015 7:99 DOI: 10.1186/s13073-015-0220-9




Greninger, AL, et al.l;*Genome Medicine* 2015 7:99 DOI: 10.1186/s13073-015-0220-9




Greninger, AL, et al.l;*Genome Medicine* 2015 7:99 DOI: 10.1186/s13073-015-0220-9




Greninger, AL, et al.l;*Genome Medicine* 2015 7:99 DOI: 10.1186/s13073-015-0220-9

- Able to identify correct strain in each sample, and the correct genome >90% of the time
- Able to do this in ~6 hours
- Quicker, cheaper, and possibly in the field

# Questions?

# In small groups

- Come up with a pro/con list for each of the technologies we discussed