

## Next Generation Sequencing

Sequencing human genome- painfully slow  
How can we speed things up?

Parallel sequencing reactions  
done at the same time

A cartoon lightbulb character with a smiling face, arms, and legs, standing next to the text.

### 454 Sequencing

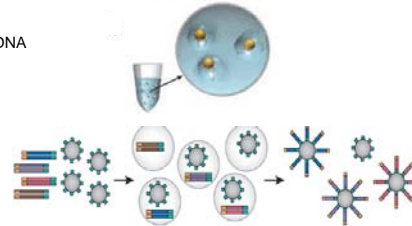
- Isolate DNA
- Amplify DNA
- Perform pyrosequencing (sequencing by synthesis)
- Detect incorporation of each nucleotide

### 454 Sequencing

Isolate DNA



Amplify DNA

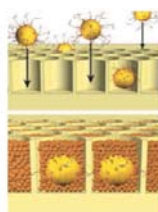
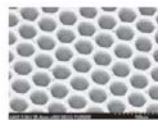


Rothberg and Leomon Nat Biotechnol. 2008

Shendure and Ji Nat Biotechnol. 2008

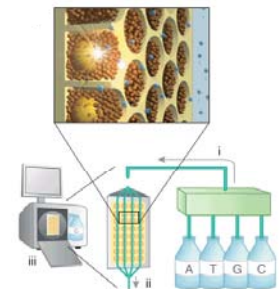
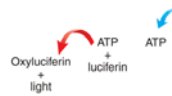
### 454 Sequencing

Perform pyrosequencing



### 454 Sequencing

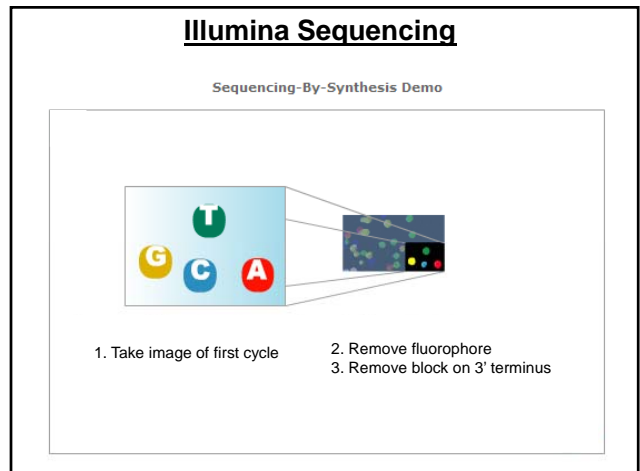
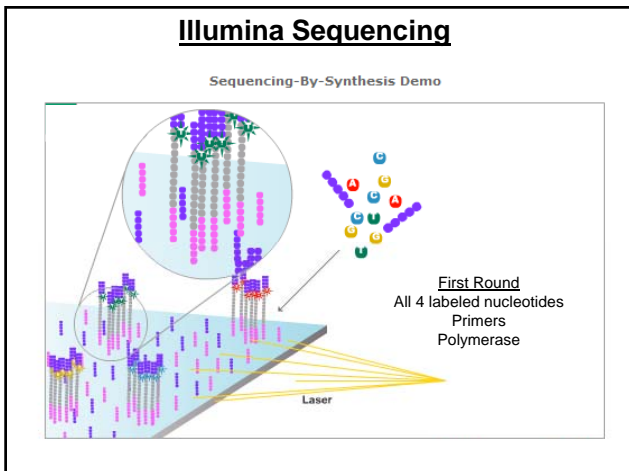
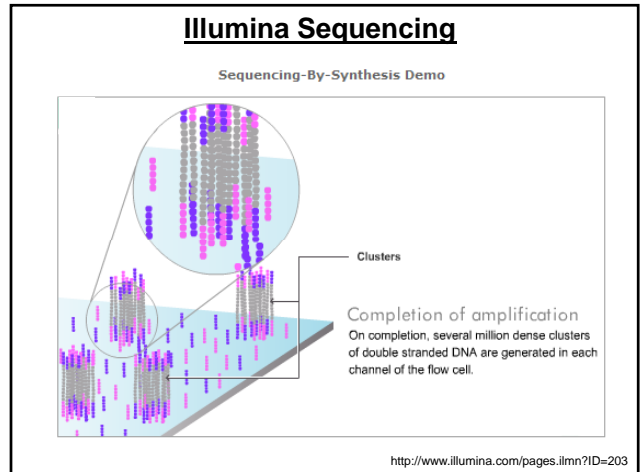
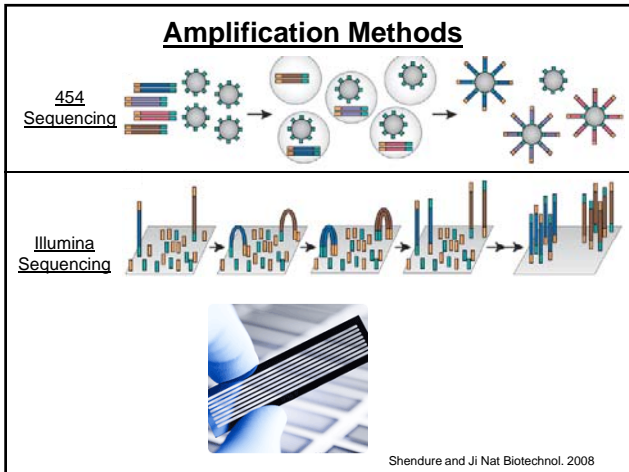
Detection of nucleotide incorporation



**Stats:**

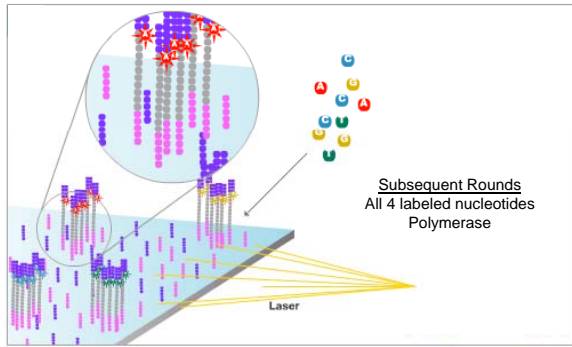
- read lengths up to 400 bp
- 400,000 reads per run
- accuracy problem with homopolymers
- costs \$60 per megabase

Rothberg and Leomon Nat Biotechnol. 2008



## ILLUMINA Sequencing

Sequencing-By-Synthesis Demo



## ILLUMINA Sequencing

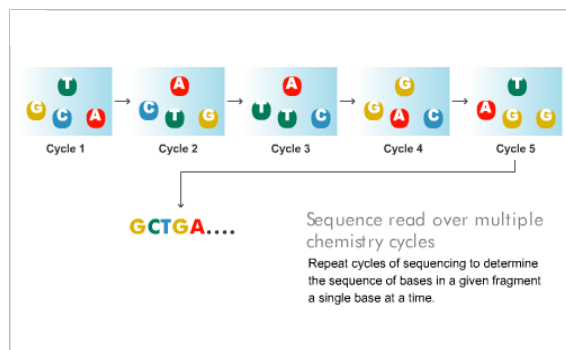
Sequencing-By-Synthesis Demo



1. Take image of next cycle
2. Remove fluorophore
3. Remove block on 3' terminus

## ILLUMINA Sequencing

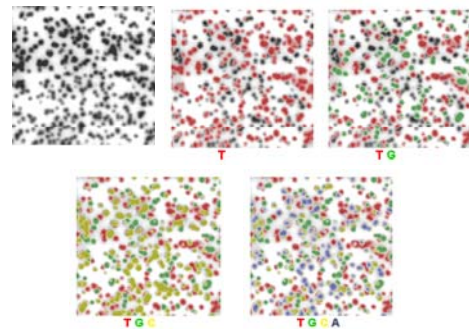
Sequencing-By-Synthesis Demo



## ILLUMINA Sequencing

### Stats:

- read lengths up to 36-75 bp
- error rates 1-1.5%
- several million "spots" per lane (8 lanes)
- cost \$2500 per lane (outside of UW)



<http://seq.molbiol.ru/>

## Illumina Sequencing Uses

- gene expression
- genomic variations
- small RNA discovery
- protein-DNA interactions

## Illumina Sequencing

### Single read

large-scale reads

Can detect:

- Expression identification
- SNP identification
- ChIP DNA identification

### Paired-end

large-scale genomic variation

Can detect:

- chromosomal rearrangements
- copy number variations
- insertions deletions (indels).

## Illumina Sequencing on Campus

### Specs

36 bp, 50 bp, 75 bp run options  
 Raw accuracy >98.5%  
 Paired-end capability  
 20-30 Gbases per run  
 2-2.5 Gbases per day  
 138-168 million reads  
 2-9 TB of information per run  
 Run time (36bp) ~2.5 days  
 $10 \times 10^6$  –  $20 \times 10^6$  reads per lane

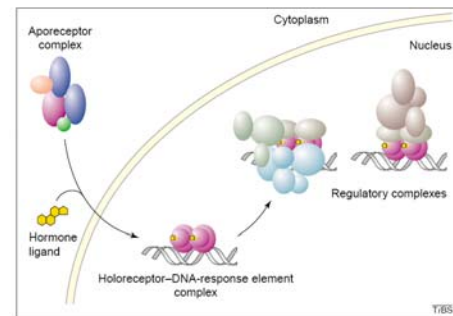
### Coverage requirements

23x coverage for SNP calling  
 15x coverage for binding (ChIP-seq)  
 7-10x coverage for de novo sequencing

Example- E coli genome  
 4.5Mb at 50 bp reads = 30x coverage  
 So: 2-3 bacterial genomes/ lane is feasible

Example- human genome  
 $10 \times 10^6$  reads to detect transcripts  
 $70-80 \times 10^6$  reads to detect SNPs

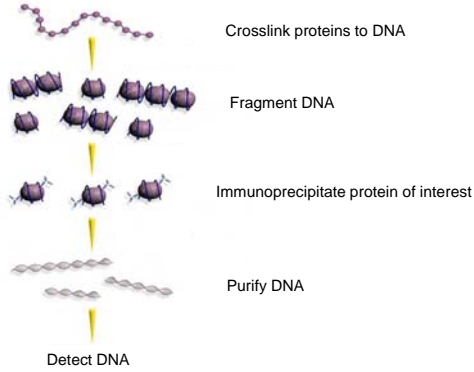
## Example- Where does estrogen receptor bind in response to different ligands?



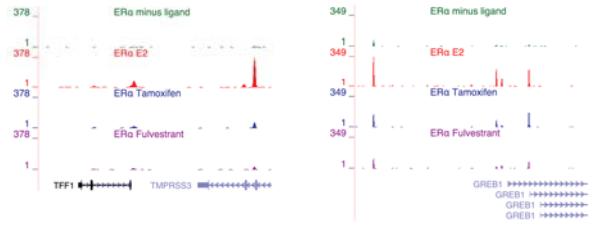
Freeman & Yamamoto (2001) Trends Biochem Sci 26:285.

## The Experiment

### Chromatin Immunoprecipitation

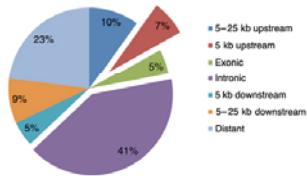


## Individual Gene Views



Welboren et. al. (2009) EMBO 28:1418

## Genome-wide comparison



Welboren et. al. (2009) EMBO 28:1418

## Resources on Campus

### Sanger sequencing

Josh Hyman  
2360 Biotechnology Center  
425 Henry Mall Madison WI 53706  
Phone (Office): 608/263-9880 (Lab): 608/263-9882

### Illumina sequencing

Sandra Splinter BonDurant  
UWBC Gene Expression Center  
890-0166

Or

Marie Adams, MS  
UW Sequencing Facility  
425 Henry Mall, Room 1250  
Madison, WI 53706  
608-262-4657

[marieadams@wisc.edu](mailto:marieadams@wisc.edu)