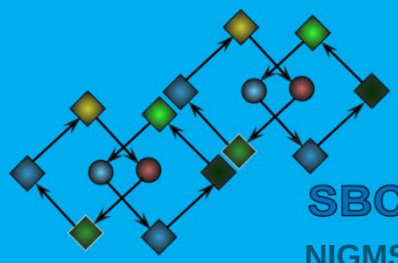


NIGMS funded Center

Experimental Methods in Systems Biology

Part of the Coursera Certificate in Systems Biology

Marc Birtwistle, PhD
Department of Pharmacology & Systems Therapeutics
Fall 2014, Week 2, Deep mRNA Sequencing

Icahn School of Medicine at Mount Sinai

Outline

- Purpose of mRNA sequencing
- Limitations of The First “Omics” Technology—Microarrays
- First, Second, and Third Generation Sequencing Technologies
- Focus on Illumina (Second generation)
  - Obtaining mRNA
  - Creating a “library”
    - Paired-end vs. single-end
    - Ligate “adaptors”
    - Amplify by PCR
  - Sequencing
  - Examples of results
- Quantification

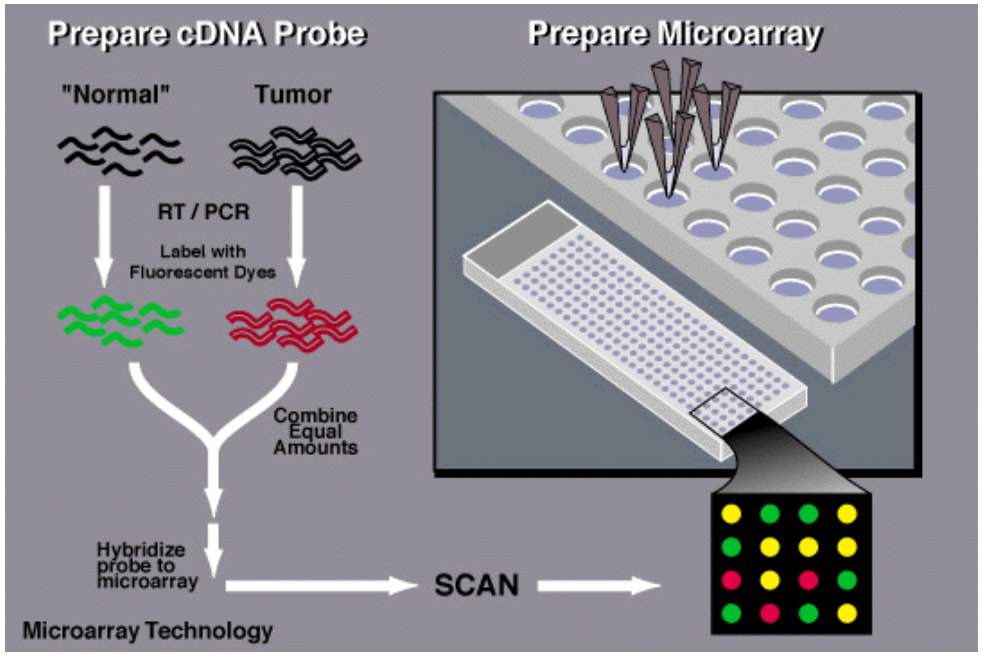
What is the Purpose of mRNAseq?

- mRNA sequencing allows:
  - Quantification of all expressed transcript levels
    - There can be issues with how quantitative it is because of PCR steps in some technologies
  - Quantification of ratios of splice variant levels
  - Identification of novel splice variants

The Big Picture: Applications of Next Generation Seq technologies

- **Impact on Research:**
- Redefines questions researchers can ask
  - From 'Does this protein regulate gene X?' to 'What genes does protein X regulate?'
  - From 'Does this gene cause disease X?' to 'What differences in the genome caused disease X?'
- New applications being developed rapidly – Plenty of opportunity to be creative!

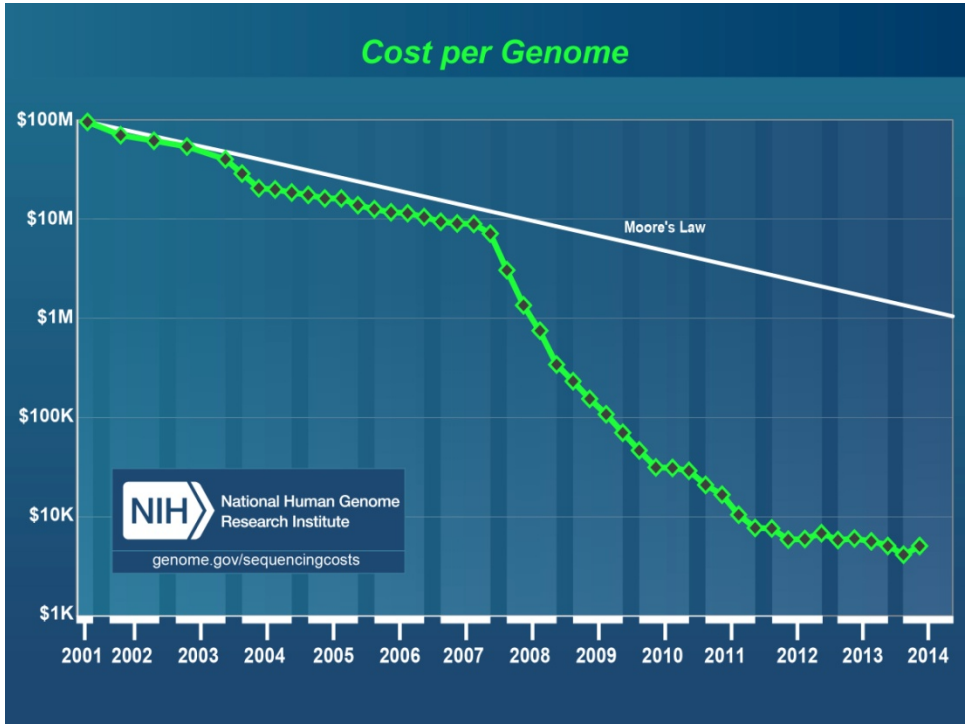
The First Transcriptomics Technology—Microarrays



http://www.genome.gov/10000533

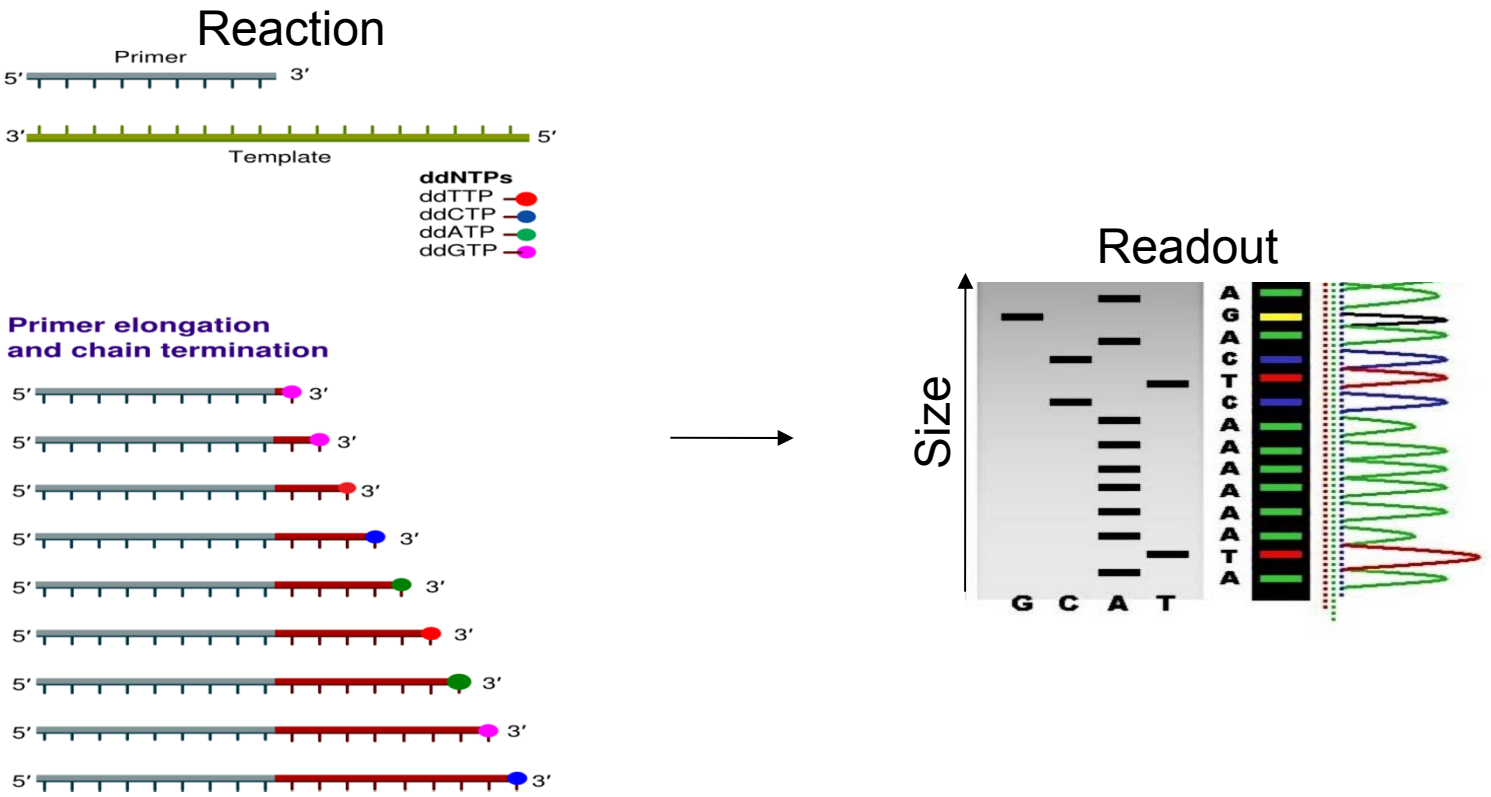
- Microarray based technologies are currently cheaper (but probably not for long)
- Disadvantages
  - Results are difficult to compare across platforms (i.e. different microarray companies)
  - You have to know what you are looking for
    - Hard to identify novel transcripts or splice variants
  - Quantifying alternative splicing is difficult (but not impossible)
- In the long run, the pure sequence data provided by mRNA sequencing will likely prove more reliable and easy to share

Evolution of Sequencing Technologies



http://www.genome.gov/images/content/cost_per_genome2.jpg

First-generation Sanger sequencing



http://en.wikipedia.org/wiki/Sanger_sequencing

Benefits and drawbacks of Sanger sequencing

Throughput:	One strand at a time
Read length:	~700 bp
Chemistry:	Chain termination
Advantage:	Highly accurate
Drawbacks:	Very slow and expensive (on a genome scale)

Increasing DNA Sequencing throughput

- Sequencing many many DNA strands at a time to increase throughput
- Automation of Sequencing process
- 96 and 384 well format for Sanger sequencing was developed.
- Fluorescent labeling of ddNTPs and capillary electrophoresis developed that enabled automation of Sanger sequencing.

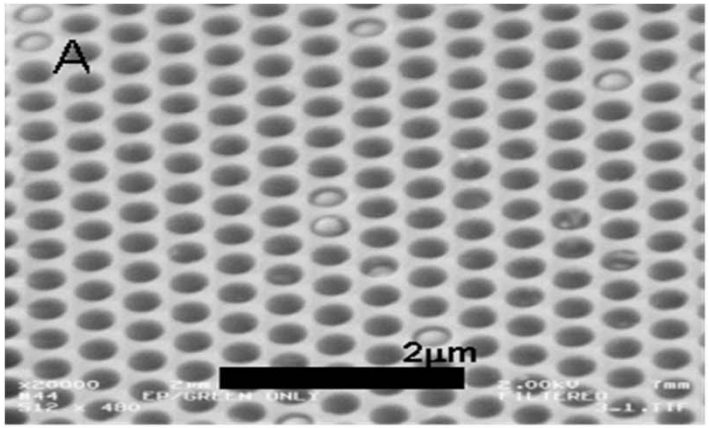
Is this increase sufficient?

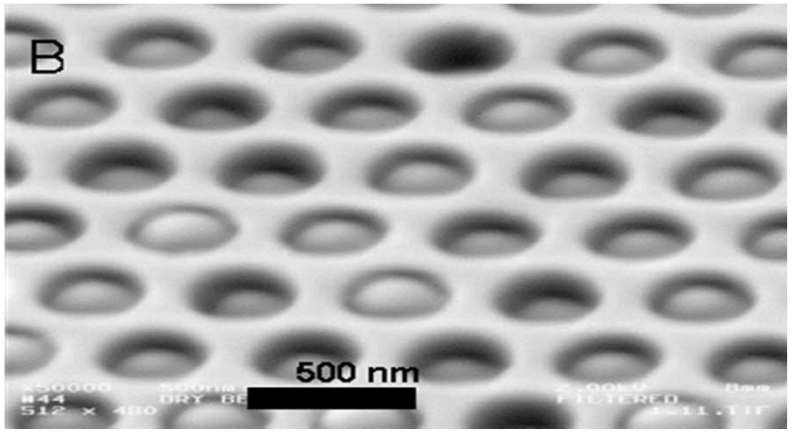
Even with increase in throughput and automation, it took 13 years to complete 1 human genome sequencing at the cost of ~$2.7 billion!

Need to increase DNA sequencing by several order of magnitude

Increase number of wells per location from existing maximum of 384 well plate

Literally go down to micro/nano well plate format….
Increase to accommodate as many as million and more wells per plate!!





Can we use dd Chain termination chemistry to sequence DNA in >million microwell format?

Sanger's dd chain termination chemistry generates several labeled DNA strands that necessitates separation of DNA strands generated in increment of one base length per Base of sequence.

So, we need to think of different sequencing chemistry that does not require DNA separation step.