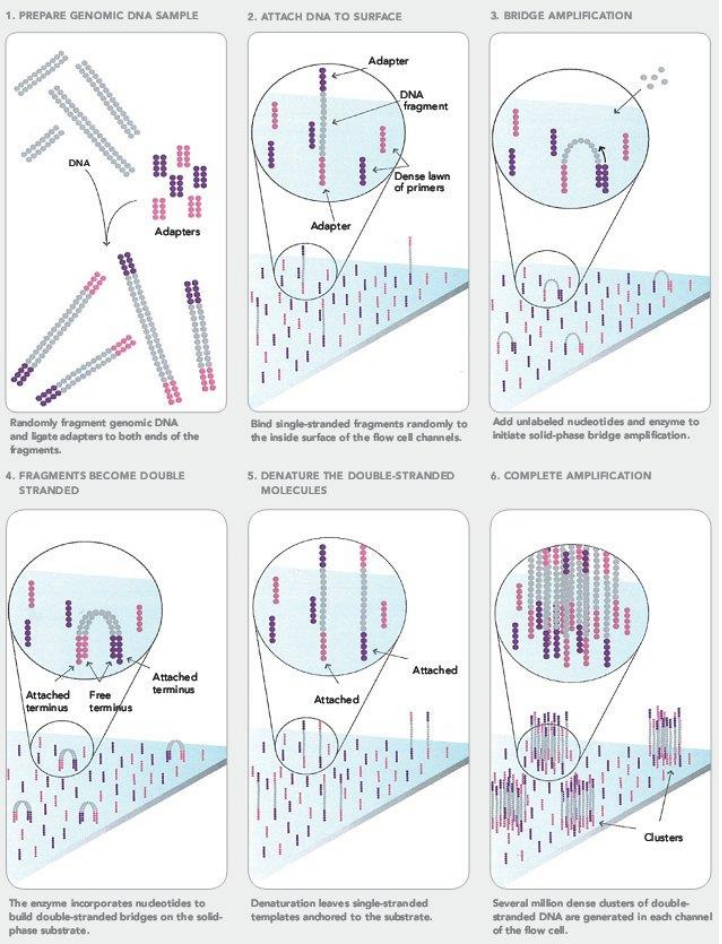
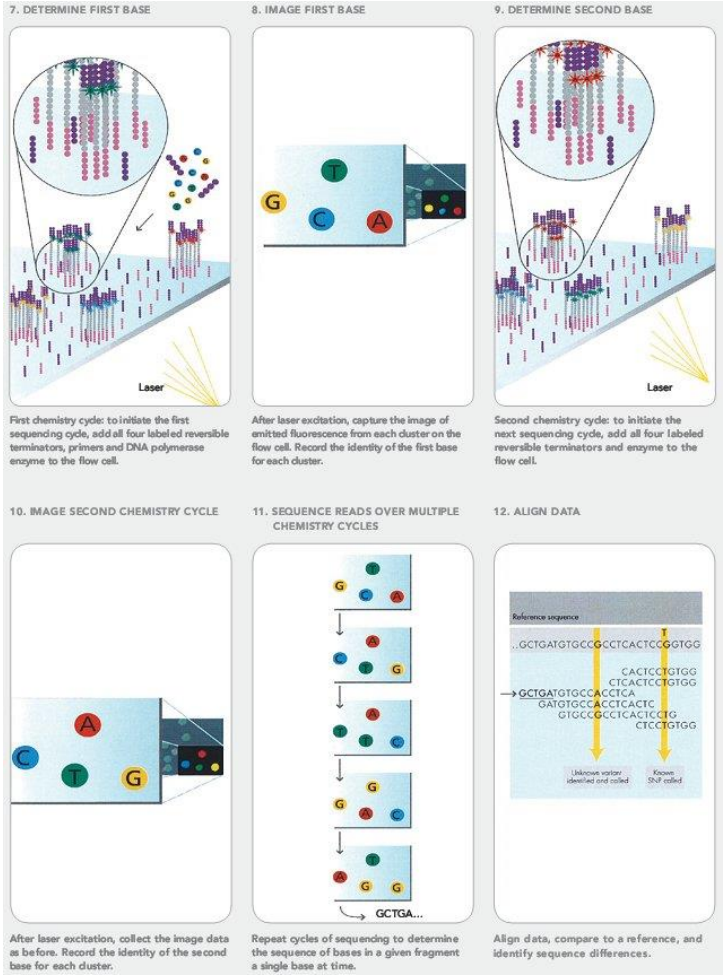
# Introduction to high-throughput sequencing with emphasis on exome sequencing

Bi188 Spring 2013. April 12th.

Georgi Marinov

# Illumina sequencing technology overview



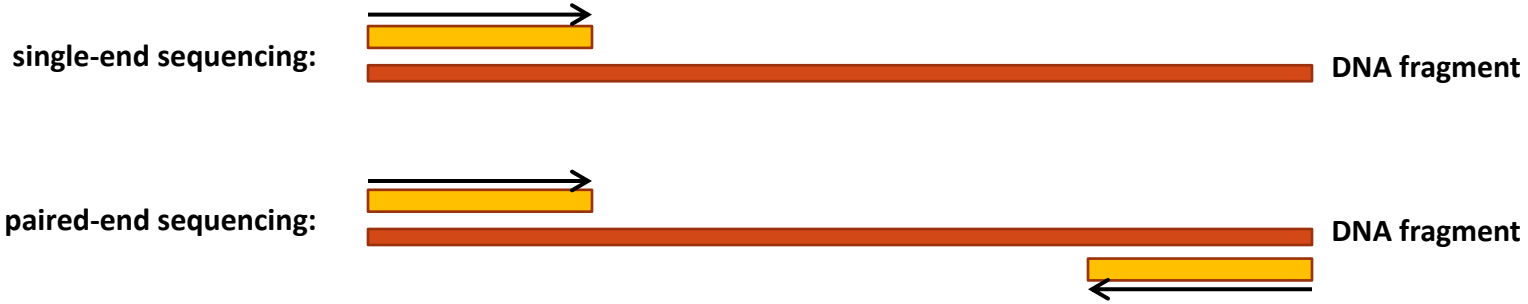
http://www.youtube.com/watch?v=l99aKKHcxC4

# Characteristics

- **Short reads. Used to be 25, then 36, then 75, then 100, now you can get 150 and even more**

- **Low error rate; errors are usually substitutions while there are a lot of indels in 454, PacBio and Ion Torrent data (and there will be in nanopores) due to homopolymer phasing errors, etc.**

- **Paired-ends:**

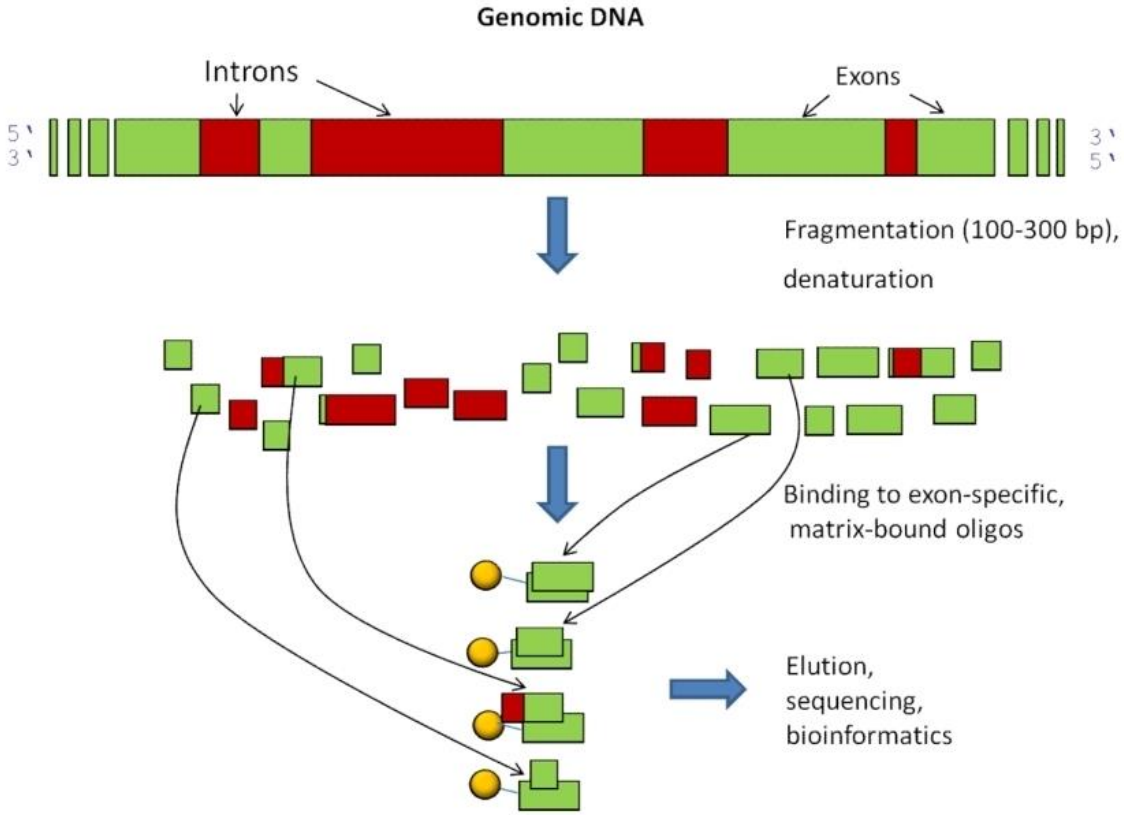



**Look up what “mate-pairs” and “jumping libraries” are if you are curious**

# Exome sequencing

**The human genome is very large and sequencing it used to cost more than $100K even with high-throughput sequencing until a few years ago. For this reason, people developed exome sequencing under the assumption that most variants with relevance to disease reside in the exons of genes. This allowed a much higher number of samples to be sequenced for the same amount of money.**

**Costs have dropped significantly since 2007-2008, but it still much cheaper to sequence an exome than to do a whole genome.**




http://www.genxpro.info/media/Whole%20Exome%20Sequencing.jpg

# Different strategies for exome capture

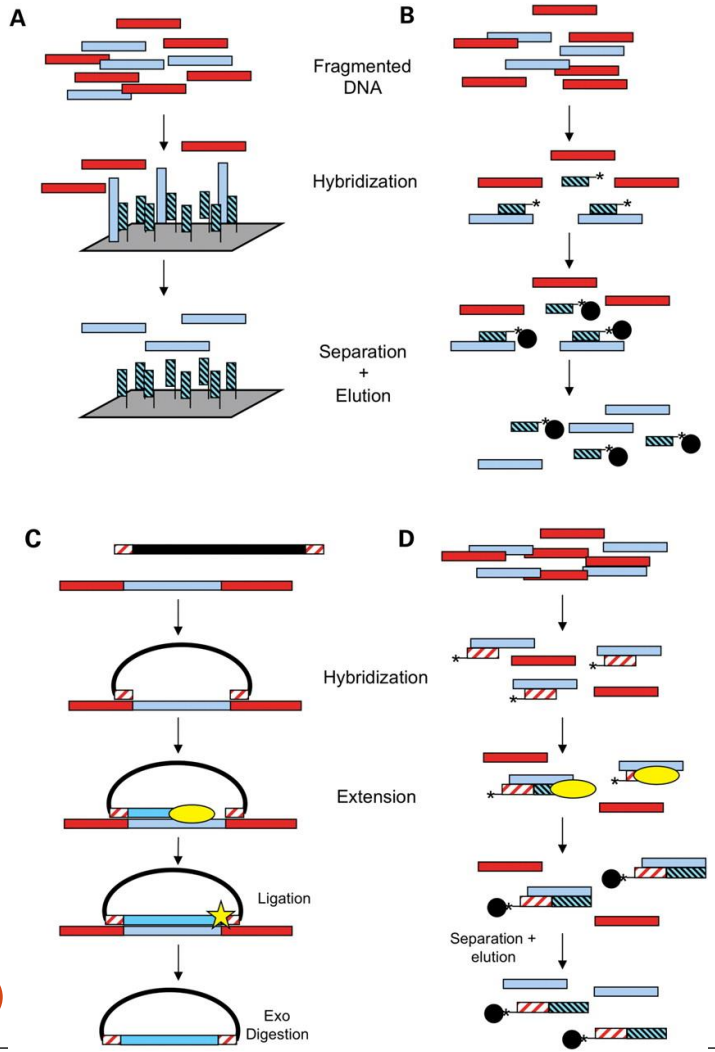



Illustration of different capture methods. Light blue bars represent desired genomic sequence, red bars represent unwanted sequence. (**A**) Solid-phase hybridization. Bait probes (light blue and black) complementary to the desired sequence are synthesized on a microarray. Fragmented genomic DNA is applied, and the desired fragments hybridize. The array is washed, and desired fragments are eluted. (**B**) Liquid-phase hybridization. Bait probes (light blue and black) complementary to the desired regions are synthesized, often using microarray technology. The probes are generally biotinylated (asterisk). The bait probes are mixed with fragmented genomic DNA, and the desired fragments hybridize to baits in solution. Streptavidin beads (black circles) are added to allow physical separation. The bead-bait complexes are washed, and desired DNA is eluted. (**C**) MIP. Single-stranded probes composed of a universal linker backbone (black line) and arms complementary to the sequence flanking desired regions (red and white) are synthesized, often using microarray or microfluidics technology. The probes are added to genomic DNA and hybridize in an inverted manner. A polymerase (yellow oval) fills in the gap between the two arms. A ligase (yellow star) seals the nick, resulting in a closed single-strand circle. Genomic DNA is digested with exonucleases, and the captured DNA is amplified using sequences in the universal backbone. (**D**) PEC. Biotinylated primers (red and white) are added to fragmented genomic DNA, where they hybridize to the desired sequence. A polymerase (yellow oval) extends the primer, creating a tighter interaction. Streptavidin beads (black circles) are added and are used to physically separate the desired DNA from the unwanted DNA. The desired DNA is then eluted.

**Teer & Mullikin, 2010**

# Data processing and analysis